

Making the market shape research

The Australian government resembles every other in wanting to turn research to profit. Its chief research agency's new policy is a step towards meeting that ambition, but is not sufficient in itself.

In relation to publicly supported research, all governments these days raise the common cry that research should have more immediate economic value. In Australia, with a population about a quarter of that of industrial economies such as those of Western Europe (which are not conspicuously successful), the cry has been amplified in the past year or so by the re-election of the Hawke government, determined (as it seems) to review and modify the working of every important public institution (see *Nature* 316, 185; 1985). The chief research organization, the Commonwealth Scientific and Industrial Research Organization (CSIRO), where pressure has been especially strong, has been a ready target for criticism. It is big, employing a substantial fraction of Australia's researchers, and is moreover still largely cast in the mould of earlier decades, when the universities of Australia were much weaker in research.

The inevitable result is that CSIRO is conspicuous among the public institutions under scrutiny; the government's chief advisory body is due to report shortly. Meanwhile, as if to show that CSIRO is not waiting, fatalistically, for whatever changes may be decreed, it has put out a document, *Shaping the Future*, which shows at least alertness to the Australian government's anxieties. The policies outlined in the new plan will undoubtedly help to forestall criticism. Thus CSIRO says it will look for ways of making its staff more flexible, and in particular to succumb more often to the temptation to work in industry. To be fair, the organization has been grappling with impediments to mobility (such as pensions schemes) for several years, without much help from government. It is, however, well known that the most effective way of "transferring technology", as the saying goes, is to move people. Published digests of "research achievements" by comparison cut very little ice. Maybe CSIRO will now get the backing it deserves; other governments may be tempted to follow suit. CSIRO also promises to be more self-conscious about economic possibilities when embarking on new research and to be more deliberate and systematic in the evaluation of research programmes during their execution. For an organization which confessed earlier this year to not having an economist on its staff, this is also a self-commendation, but one whose weight can be determined only by performance.

The much more serious difficulty is to strike the right balance between the different motivations for industrial innovation and development. What attention should, for example, be paid to industrial companies' assessments of market needs and what to economists' long-term assessments of the pattern of industry? To what extent should a government build mechanisms for seeing that new ideas arising from basic research are indigenously exploited, when it might be easier and more profitable to trade the patent rights? How far should a government go in supporting research intended to strengthen a traditional industry (agriculture, for example, in Australia) and what resources should it spend on seeking a toehold in high-cost high-tech industries in which every other country trumpets similar ambitions but where only some can hope to succeed? CSIRO will no doubt argue behind the scenes that all these questions, nowhere easy, are complicated in Australia by the relics of the industrial strategies of past decades that sought to encourage the growth of Australian industry by constructing shields against outside competition. Is it any wonder that one of the present government's discontents is that commercial companies in Australia spend too little on research? The

obvious snag is that no amount of cleverness in CSIRO laboratories, and no amount of persuasiveness in publicizing the innovations that result, will counteract such disincentives to innovation. The new proposals for counting industrial research spending as a tax-deductible expense worth half as much again may help, but letting Australian industry sense how competition strengthens the need for research would be even more effective. Whether the reforming government will go so far is another question.

Australia's special difficulties are to some extent historical, stemming from CSIRO's past role as virtually the only agency of public research. Everybody will appreciate how difficult it must be for an organization deliberately to relinquish responsibilities it discharges well. Fortunately, there are signs in this brief strategy document that CSIRO is now leaning in the right direction. What remains to be seen is whether the new management in Canberra will be able to turn these general principles into a more pointed pattern of research — and whether the government will allow it the time to do so. □

Fair play for foreigners

Japan is committed to internationalization. But recent events suggest it must look again.

SOME three years ago the Japanese government altered the law in order to permit foreigners to hold permanent posts in the state universities. The change was made partly because there was an awareness in the academic community that Japan was isolated from the tradition of employing talented staff, regardless of nationality, which has done much to enhance the vitality of European and American universities. At the same time the change did away with the difficulties faced by Koreans who had been born in Japan or who had come to Japan during the war.

That Japanese scientists are aware of the benefits of such "internationalization" is clear enough, for a growing number now hold tenured posts in foreign lands. But despite the hopes held out by the change in the law the complementary process — that of giving foreigners full-time employment in Japanese universities — has not made much headway.

In part this reflects language difficulties and the rather rigid seniority system of the state universities which provides few cracks through which those outside might enter. Only one university — Tsukuba University — has done away with the university chair system and, by no coincidence, it also has by far the largest number of foreign staff members. But as events described on p.465 show, problems still remain at Tsukuba. It would seem that some foreign teachers' legitimate concerns for their future security have been thought little of at the university, while other foreigners find they are welcome so long as they remain only for a short period. To the small band of foreigners who work in Japan's state universities these events are important, for Tsukuba has long been seen as the leading proponent of the internationalization of the universities.

The spectrum of attitudes seen at Tsukuba seem well to reflect the deep-seated ambivalence the Japanese themselves feel towards "internationalization". For some years the internationalization of Japanese life has been a topic of endless discussion, although there is far from agreement on what it means. It is



clear, though, that more than a century after Japan was opened to trade, differences in language and culture are still felt to provide barriers to intercourse with the rest of the world.

From the inside, Japan can seem a small, homogeneous, hierarchical and conservative country far removed from the vaster flux of events occurring in the United States and Europe. The distant West often seems romantic — a feeling Japanese TV commercials have learnt to exploit with great skill.

But the West can be seen as threatening, too, with cultural and language barriers providing convenient protection for Japan. After all, will not “internationalization” mean “westernization” and risk the destruction of the form of society that has reaped such tremendous material benefits for Japan?

In public life the oscillations between the view that Japan should become part of the larger world and the view that its uniqueness must be protected seem to continue endlessly. Recent weeks have seen a sudden insistence by the Ministry of Education, Culture and Science that the national anthem be sung and the Rising Sun flag be hoisted at school ceremonies in an attempt “to strengthen national identity”.

The same aims can be seen at work in proposals to set up an “International Centre for the Study of Japanese Culture” in Kyoto. When the idea was first mooted it was for an institute where foreign scholars with an interest in Japan could carry out research and have effective access to source materials. But recently Prime Minister Yasuhiro Nakasone has taken a strong interest in the centre and its stated goal has become “to explain, defend and export Japanese, cultural values to the rest of the world”. The Prime Minister is quoted as saying that “Forty years have passed since the end of the war . . . now is the time to establish Japan’s identity once again. And that is why I am calling for the creation of the centre”. Among those tipped to act as advisers to the centre is Kinji Imanishi, a man often described as a uniquely Japanese thinker, but whose anti-darwinian evolutionary theories, inspired by his view of Japanese society, would not last long if subject to international attention.

Despite such pronouncements, Nakasone is most vigorous in his calls for the internationalization of Japan. But his view seems to be that strengthening Japan’s identity is the only way for Japan to take an equal place in the community of nations. Paradoxically, nationalism then becomes a prerequisite of internationalism. If this were so, true exchange of ideas would become impossible, for contact with foreigners would simply be a way of reasserting Japan’s “unique cultural identity”.

At Tsukuba (and at universities elsewhere) many foreigners complain of the separate treatment they receive: “Foreigners are put in a box,” as one lecturer put it, “where Japanese can come and get a controlled dose of internationalization.” But internationalization can be something more than that. At its simplest it surely means no more than just giving foreigners equal treatment and allowing them the same role in university life as Japanese teachers. There is really nothing very difficult about it: private universities have been giving equal treatment to foreign staff for years. That the national universities were originally set up to import learning to aid Japan’s modernization and that their staff have the status of civil servants should not be a barrier.

Special procedures may well be necessary to recruit foreigners, for they will not have the network of supporters usually so necessary to obtain employment. But beyond that why treat foreigners differently? Why not let the job go to the most talented person applying for it? So far, a mere four state universities have given long-term employment to just five foreigners.

Those who fear that foreigners will never understand Japanese concepts of harmony and will disrupt proceedings if allowed to serve on committees may find themselves in for a pleasant surprise. And those like Nakasone who worry about national identity can be sure that exposure to foreigners and foreign ideas will simply give further rein to the Japanese genius for absorbing the best ideas from anywhere they can be found. Nothing is more likely to kill Japaneseness than a deliberate attempt to inculcate it. □

Is freedom a luxury?

The issue of tenure threatens British academic peace. But there is a new plan for change.

THE doctrine that those who work in academic institutions should be entitled to life-long tenure of their posts has become a divisive issue between academics and the rest of society, especially in financially hard-pressed countries such as Britain, where the government has embarked on a legal procedure to extinguish such right of tenure as there may be in contracts of employment issued by universities to the members of their teaching staffs. Since the chief consequences of that procedure will be a worsening of already strained relations between the government and the universities, those concerned should be compelled to read the interesting document on the subject put out last week by the Council for Science and Society (*Academic Tenure — Luxury or Necessity*, £2.50, from 3–4, St Andrew’s Hill, London EC4). The document has been drawn up by a committee under Sir Stephen Bragg. Properly, it points to the general misconception, especially among British politicians, about the degree to which academics are protected from misfortune by their contracts of employment; academic life is not really as secure as generally supposed. Rightly, the document also argues that the social function of academic institutions would be compromised if academics could be sacked as easily as if they were workers at, say, a building site, which would often be a recipe for suppressing even constructive unorthodoxy, but it goes on to ask that academics themselves should appreciate the difficulties that must arise when universities are forced, by financial pressures of the changing pattern of student demand, to think of closing whole departments. These days, circumstances can arise when tenure rights become threats to the survival of whole institutions.

The particular value of the document, however, is that it carries the argument a step further than most other discussions of the right of tenure, arguing that the right of academic tenure cannot be considered in isolation. In Britain, the circumstances are complicated by the way in which salaries are determined; young people joining the academic profession win the right to tenure only after a probationary period, but then are placed near the bottom of a salary ladder which assures them of annual increments (supplemented, as a consequence of national negotiations with the universities, by upward adjustments for inflation). At present, British academics may be within their rights to protest that even with these benefits, they are badly paid. The council’s report, however, breaks new ground with the assertion that the right of tenure, necessary as a means of ensuring people’s freedom of enquiry, should not necessarily include the right to a steadily increasing salary. Academic freedom would be safeguarded, the report says, if people at odds with their departments or institutions were assured only of some basic salary. To make such a system work without too much rancour, it would probably be necessary that the funds now dispensed as salaries should be earmarked as supplements paid for identifiable services.

This proposal, cautiously described as a “tentative solution”, will not readily win converts among British academics, and will have the Association of University Teachers up in arms. But the proposal deserves a sympathetic hearing, one entirely divorced from the question of what the basic salary should be (or whether present starting salaries are too low). The side effects of a change in this direction also would include other benefits, not least (under present circumstances in Britain) the increased freedom of universities to spend resources flexibly. Properly handled, the concept that academics joining an institution are paid a basic salary, and extra only to the extent that they contribute to their institution’s economic activities, could be dignifying of the academic profession and not demeaning. So, too, would be a break with the present British convention that all academics should be paid the same, however successful (or unsuccessful) their institutions may be. □

Tsukuba University

Turmoil over treatment of foreign staff

Tokyo

INTERNATIONALIZATION is a word very much in vogue in Japan; whether it refers to the yen and money markets or to the emergence of a new kind of Japanese who is equally at ease drinking cocktails in New York or *saki* in Tokyo. Internationalization of university education too is also much talked of. But recent events at Tsukuba University suggest that there is some rough going ahead.

It is claimed that four foreigners at the university, who have lost their jobs, have been shamefully discriminated against by university authorities. One of them, a Korean national, Professor Dong-jin Kang, has taken the ultimate step of suing the university. The first major court hearing will take place next week but a final verdict is not expected for two years.

Paradoxically, the problems experienced by the four foreigners — a West German, a Taiwanese, a Korean and an American — stem from changes in the law enacted in 1982 which finally made it possible for foreigners to be employed as tenured staff at state universities. Taking advantage of the new law Tsukuba University decided to give tenure to some foreigners. But it was also decided that henceforth no foreigner without tenure — that is thirty or so foreigners on contracts renewed each year — could stay for more than four years. Foreigners who had already been there longer than that were thus in effect dismissed. But four of them — all with exceptional academic records — were, they believe, repeatedly assured by senior university staff that they were to be given full tenure when their one-year contracts expired. They were told not to look for work elsewhere. But when the contracts ran out they did not receive the expected job nor a letter of explanation, simply a form of application for severance pay that signalled that they were now unemployed.

Among the four foreigners Professor Kang appears to have received the most frequent and definite assurances of continued employment. He now sits in his office — which he occupies only because of his success in gaining an injunction to prevent the university throwing him out — looking angrily through the many pages of detailed notes he took of the promises given at interviews with senior university staff. Why, he asks, for what purpose, was he so cruelly deceived? His academic reputation is considerable — he has published 15 books and more than a hundred articles — and he would have been able to find work elsewhere if given honest warning. He is a very angry man. Previously he

had “never thought of anything but his own academic work”, but now he is determined to fight, because, as he puts it “as a human being I cannot permit people to take my hand, shake it and lie to me”. The reference is to a meeting he had just a few weeks before he lost his job at which he claims two of the university’s five vice-presidents promised that all was well.

Professor Margarete Sawada — a West German with a Japanese husband — tells a similar tale. Other members of staff corroborate much of what she and Professor Kang say. And that both professors’ names are on this year’s lecture schedule suggests that many thought their continued presence certain. Professor Sawada is now unemployed, but is not taking legal action.

The president of the university, Professor Nobiyuki Fukuda, a distinguished physicist who spent some years in the United States, and a fluent English speaker well known for his international views sees recent events differently. He is not himself directly responsible for appointments: indeed, the Tsukuba constitution specifically prevents him from being so. He plainly regards it as extremely strange that any foreigners should have felt assured of permanent employment. He explains that a new post can only be created after a specialist committee has shown the need for a new post; candidates have to be interviewed before a selection committee can do its work. And even at that late stage all is not certain — if the budget does not come through then the appointment may still be cancelled. The president is adamant that Professor Kang’s appointment has never been discussed at a relevant university meeting he has attended. “Powerful professors may have given affirmative responses” to Kang’s enquiries, he says, but that just means he had “supporters but not a job”.

Has there then been a simple misunderstanding? Many university staff, foreign and Japanese, think not. Professor Fukuda’s term of office at Tsukuba is nearly over and soon elections will be held for a successor. It is a common view that it is the proximity of this election that accounts for recent events. The university has a reputation for in-fighting, perhaps because its progressive new constitution did away with the “chair” and “faculty” system and thus removed some of the time-honoured ways of dealing with disputes. Some feel that the allocation of university posts has now become too important a weapon in university power struggles for mere foreigners to be any longer considered for them. One vice-president who supported

the view that foreigners fit for full-time posts should get them and who had fought for tenure for the foreigners was fired from his post as vice-president.

Foreign lecturers also point to other discouraging events. Some who were once told they were to be treated in “just the same way as Japanese” find they are no longer welcome at council meetings, even as non-voting observers, nor are they now permitted officially to be responsible for supervising the thesis work of final year students. And another vice-president was recently reported as saying that “foreigners should not teach something a Japanese could teach and that long-term employment should only be bestowed on foreigners who are Nobel prize candidates”. Small wonder then that many foreign teachers are wondering whether they really count for much at Tsukuba.

The rule limiting employment to four years is seen as particularly repressive. Its justification is that it is better to have a large number of different teachers than a few who stay a long time. But if that is true why not insist all staff — Japanese and foreign alike — are replaced every four years? Clearly the policy would be disastrous. The rule only makes sense to someone who sees foreigners as different species from the Japanese; not to be treated as equal and responsible members of the university staff.

The rule also ignores the emergence of a quite new breed of foreign teacher. Not so long ago many foreign teachers in Japan had no intention of remaining for any length of time. But now graduates with a particular academic interest in Japan’s language and culture, plus an ability to speak Japanese and as often as not a Japanese spouse, are common. These people are willing to give the commitment that a good university needs.

It is a pity that the atmosphere at Tsukuba should have been soured in this way for it is by far the most international of the state universities. Five per cent of its students are foreign, there are more than thirty foreign lecturers on short-term contracts and many facilities are specially designed to make it easy for foreigners to come and study. Nothing like this has been achieved anywhere else and there is much to be proud of. To Professor Fukuda, the view that Tsukuba is becoming more parochial and xenophobic is “non-sense”. And he firmly proclaims that “internationalization” is Tsukuba’s greatest goal. But curiously just as the news of goings on at Tsukuba began to reach the press, an article appeared proudly stating how Prince Hiro, the Emperor’s grandson, who has spent two years doing research at Oxford University, has been treated “just like anyone else” during his stay. If that seems close to the true meaning of “internationalization” then some further thought is needed about the direction Tsukuba University is taking.

Alun Anderson

AIDS

More for research and treatment

Washington

MOUNTING concern over the epidemic of AIDS (acquired immune deficiency syndrome) has prompted Congress to add large sums to the administration's proposed budget for combatting the disease in fiscal year 1986, which started on 1 October. The House of Representatives has already passed an appropriation bill that would allocate \$189.7 million for research, prevention and treatment, \$70 million more than the administration's request and double what was spent last year, and the Senate appropriations committee last week voted to allocate a total of \$219 million to the disease.

Much of the extra money for AIDS would go to the National Institutes of Health (NIH); the House has earmarked all of its \$70 million increase for NIH, whereas the Senate committee spread its bounty between NIH (\$128 million), the Center for Disease Control (\$62.9 million) and the Alcohol, Drugs and Mental Health Administration (\$12.7 million). The \$16 million remainder in the Senate's allocation was voted to support a special fund to be administered by the Secretary of Health and Human Services for demonstrating AIDS treatments. The Senate bill has yet to pass the floor, but is not expected to be challenged on its provision for AIDS and the small differences with the House will then be settled.

The total appropriation for NIH in fiscal year 1986 is likely to come out at about

\$5,460 million, \$810 million more than the President's request and \$1,000 million more than the budget for the current fiscal year. This assumes no major attempt to make cuts on the Senate floor. The number of extramural research grants agreed by the Senate committee for 1986 — recently a bone of contention with the administration — is 6,000 the same as the number of multi-year competitive grants this year.

Tim Beardsley

• A US Public Health Service (PHS) plan for prevention and treatment of AIDS published recently is markedly more pessimistic in its assessment of prospects for a vaccine or therapy than was ex-Secretary of Health Margaret Heckler last year. In contrast to Heckler's prediction that a vaccine would be available within two or three years, the PHS plan concedes that "it is unlikely that a vaccine to substantially limit transmission will be generally available before 1990".

The plan sets goals of reducing the increase in transmission of AIDS virus by 1987; of reducing the increase in disease incidence by 1990; and of eliminating transmission of the virus by the year 2000. Edward Brandt Jr, the former assistant Secretary of Health who organized the PHS executive task force on AIDS, defended the goals as realistic but admitted to some doubts over whether the goal for the year 2000 would be totally achievable. Nevertheless, he said, progress indicates that "we're on the right track". □

Nuclear winter

US arms control policy doubts

Washington

THE US Senate Armed Services Committee heard for itself last week the evidence for the global climate effects of a nuclear war and considered policy changes that could arise as a result of this knowledge. Although the committee found the evidence convincing, it remained in doubt that any changes in current official arms control policy are needed.

Richard Perle of the Department of Defense (DoD) said that the administration is "indebted" to scientists for their research but their suggestions of action based on the fear of a "nuclear winter" are inappropriate. In a sometimes heated exchange with Carl Sagan of Cornell Uni-



versity, Perle admitted that the idea of a nuclear winter has caused "a lot of concern", but only provides support for the administration's existing policy of bilateral strategic arms reduction combined with increased defensive programmes (the Strategic Defense Initiative) and improved "verification technologies" to allow accurate estimation of global nuclear forces.

Sagan argued that the number of US strategic nuclear weapons has, in fact, been increasing each year since the mid-1960s, stressing to senators that as few as 100 nuclear explosions would have severe climatic consequences. Senator Barry Goldwater (Republican, Arizona), chairman of the committee, pointed out that should a nuclear winter scenario be realistic, the greater accuracy of US weapons (current DoD thinking favours smaller and more accurate missiles) becomes an irrelevant issue.

Perle admitted the need for DoD to spend money to assess more accurately a lower limit for the number and magnitude of nuclear explosions that would affect the climate. But he saw no need for research on the biological effects of nuclear war at a time when the physical effects are so uncertain. And Sagan's conclusions that "drastic" reductions in nuclear arsenals are called for drew little response from the committee.

Maxine Clarke

New US policy on data security

Washington

THE White House has announced an administration-wide policy of relying exclusively on national security classification as a means of controlling access to fundamental research data. Universities and laboratories have been waiting anxiously since the proposal was first announced in May 1984 by Edith Martin, then Deputy Under-Secretary of Defense for Research and Engineering.

The Department of Defense (DoD) already has in place a policy of placing no restrictions other than security classification on fundamental research data; nevertheless, the 16-month delay in making the policy general is widely attributed to opposition within certain factions of DoD — notably the international security policy division — to any loosening of DoD's powers to restrict the flow of research data to the Soviet bloc. The new policy takes the form of National Security Decision Directive 189. It declares that "to the maximum extent possible" the products of fundamental research are to remain unrestricted; where controls are necessary, classification is to be the applicable technique,

with each government agency responsible for determining whether to classify its research projects before making awards and to review periodically the status of all their research grants.

As most universities do not allow classified research on campus, the new policy should ensure that secrecy controls are not applied retrospectively to campus research.

Nevertheless, the victory for the universities may be largely symbolic, since few agencies other than DoD support much unclassified fundamental research that might have been restricted: administrators at the Department of Energy, for example, are said to be wondering what they should do in response to the directive.

Robert Rosenzweig, president of the Association of American Universities, cautions that the directive does not address one of the universities' current major concerns: DoD's policy of applying export controls to sensitive but unclassified data arising from research that is not considered to be fundamental. On this topic there is "still a dialogue", according to Rosenzweig.

Tim Beardsley

Disease research

Tropics still a low priority

Washington

THE small amount of money spent by the United States on tropical disease research (\$100 million of an annual biomedical research budget of \$4,000 million) reflects policy choices and not the lack of promising avenues of research, according to an Office of Technical Assessment (OTA) report published last week. Much of the recent progress has been made by small but "highly motivated" groups of researchers. Biotechnology and new techniques in immunology, combined with traditional parasitology and infectious diseases research, make the possibilities for alleviating the suffering caused by tropical diseases greater than ever before, OTA says.

The report stresses the "enormous potential" of genetically engineered vaccines, although the first of these, the malaria vaccine, is unlikely to be generally available for 5–10 years. No vaccines yet exist for human parasitic diseases and there are only a few for viral and bacterial diseases; biotechnology will allow safe and effective vaccines to be developed for many of these diseases.

OTA says that insecticides are still useful in disease control, although no single technology is likely to be able to control most vectors of disease. Integrated pest management, a relatively new approach that combines biological and chemical methods, has not had time to prove itself, but OTA says that long-term solutions may be provided by these methods.

OTA's problem is how to encourage research in an area that has poor commercial rewards. It cites some successes, for example, praziquantel, a drug that has revolutionized schistosomiasis treatment but is expensive; oral rehydration therapy

for diarrhoeal diseases; and ribavirin, a drug that seems to be effective against Lassa fever in West Africa. But federal science budgets are tight and have other priorities; pharmaceutical companies are reluctant to invest in the development of drugs of no commercial value.

One suggestion made by OTA is that tropical disease drugs be classified as "orphan" drugs, which would provide certain statutory benefits to companies developing them. The government could also guarantee the purchase of products and assistance in field trials, or authorize funds to establish a nonprofit organization to undertake research until the products are attractive enough for the private sector to take over. In the short term, however, the report suggests that a special congressional appropriations hearing be held to take evidence from those organizations that support tropical disease research and to identify areas where effort could be increased. An increase in federal funding or the removal of statutory limitations on the amount of research supported by the Department of Health and Human Services might result from such an investigation.

For the 5 million US citizens estimated to be at risk of contracting tropical diseases, many of them in the armed forces, and the millions of people in the developing countries who die as a result of these diseases each year, there is no immediate chance that the United States will channel more resources towards research in this field.

Maxine Clarke

"Status of Biomedical Research and Related Technology for Tropical Diseases" (OTA-H-258) is available from the US Government Printing Office, Washington, DC, price \$11.00.

The world's legumes go on-line

WITH the inauguration of an ambitious new database and information service, anybody with access to a personal computer should be able to tune in to the world's expertise on the 17,000 species in the Leguminosae family.

The legume family includes pulses such as peas, beans, lentils and soya, forages such as clover and vetches, sources of insecticides and haemagglutinins, and even trees for timber production. To handle the data generated by such diversity, the "ILDIS" database — the International Legume Database and Information Service — was established at a workshop held at Southampton University last month. The project combines the data collected by 16 research groups, with major contributions (in funds and facilities) from the Royal Botanic Gardens Kew, Missouri Botanical Garden St Louis, the University of Southampton, Biosciences Information Service

(BIOSIS) of Philadelphia and the UK Science and Engineering Research Council.

The database is to include details of the taxonomy and chemistry of each plant, distribution, potential agricultural uses and conservation status. The information is to be available online from Philadelphia in the United States and from York in the United Kingdom. To ensure a wide availability for the ILDIS data — to the arid and humid tropical areas where legumes are the major crop for instance — it will also be available on standardized computer software that can be provided by post.

For more information on ILDIS, potential users and contributors should contact the project coordinator, Dr Frank Bisby at the Biology Department, Southampton University, Southampton SO9 5NH, UK. The next challenge will be the largest group of plants — the Triticales, including the grasses and cereals.

Charles Wenz

Soviet psychiatry

Aiming for rehabilitation?

THE Soviet All-Union Society of Neurologists and Psychiatrists seems to be angling for readmission to the World Psychiatric Association (WPA). The first ever congress of psychiatrists from socialist countries, held in Moscow last month, was used by several Soviet psychiatrists and their Comecon colleagues to drop appropriate hints. The Soviet Union resigned from WPA in January 1983, preempting a motion of exclusion tabled for the World Psychiatric Congress in Vienna the following July. Subsequently, the Cuban, Bulgarian and Czechoslovak associations resigned also.

The motion of expulsion was a reaction to a growing body of evidence on the Soviet misuse of psychiatric methods and psychotropic drugs for the repression of political dissidents and religious believers. Last month's Moscow congress was used by the participants to deny these reports. WPA was said to have launched a "campaign of slander" against the Soviet Union and its allies. According to a Czech psychiatrist, Dr Jozef Pogady, "manifestations of disease which are clear to specialists were withheld, distorted and presented in the light needed by the organizers of the campaign" to create an "appalling myth". He was, presumably, referring to "creeping schizophrenia", a disease unrecognized in the West, but which, according to its "discoverer", Professor Aleksandr Snezhnevskii, exhibits no symptoms apart from political or religious or nonconformity. In view of this major difference in outlook, dialogue between the two psychiatric "camps" would seem futile, and the Moscow congress may well turn out to be a first step in establishing a Socialist Psychiatric Association.

For the moment, however, the Soviet psychiatrists seem more keen to reestablish their reputation with WPA than to set up an alternative. Accordingly, during the recent congress, the deputy director of the Serbskii All-Union Research Institute for General and Forensic Psychiatry (a source of many diagnoses of "creeping schizophrenia"), stated that Soviet psychiatry stands "in defence of human rights", while Dr Johannes Neumann of the German Democratic Republic claimed that since the 1983 World Congress, WPA's leadership had become "more fair and objective", and that an increasing number of member countries of WPA rejected the allegations of Soviet psychiatric abuse.

Although normally considered a hard-line country, East Germany did not withdraw from WPA in 1983 and Dr Neumann is now a WPA vice-president. He would therefore be an obvious intermediary if the Soviet Union were to seek readmission to WPA.

Vera Rich

French budget

Science spending still in vogue

FRENCH scientists are to benefit again from the strong political support given to research in France since the present socialist government came to power in 1981 — but some will benefit a good deal more than others. That is the message of the 1986 government budget announced by Prime Minister Laurent Fabius late last month, a budget in which real civil research funding will rise by around 8 per cent, while the total government budget will remain constant or even decrease. The government's target of 3 per cent of gross national product (GNP) for the national research and development budget will be reached "without doubt" by the end of the decade, research minister Hubert Curien has buoyantly predicted. The present figure is 2.3 per cent, and Curien predicts an increase to 2.4 per cent next year.

However, these and similar figures in the past, have hidden a number of hard truths. On the whole, basic research has fared less well politically than applied, but in applied sciences the new spending has not always been well-conceived; and a

considerable fall in the value of the franc over recent years has increased the real cost in France of many research materials and equipment, most of which have to be bought outside in hard-currency countries such as the United States, West Germany and Switzerland. Molecular biologists, in particular, even those in leading laboratories, consider that their budgets have only kept pace with the rising costs of research. "But that's better than many of our colleagues abroad", said one.

Nevertheless, 1986 may be different. Curien has taken on board the need for the "re-equipment" of many French laboratories and set aside a special budget for the purpose. At the Centre National de la Recherche Scientifique (CNRS) — France's leading basic research council — for example, next year's spending will include a 25 per cent increase in the budget for "medium heavy equipment", CNRS director Pierre Papon promises. "This means things like nuclear magnetic resonance spectrometers, infrared spectrometers and minicomputers, things costing FF 1–2 million (£100,000–200,000)", said

Papon. CNRS spent FF 100 million (£10 million) on such items last year, and will spend FF 125 million in 1986. Materials science, robotics, electronics and chemistry may see the greatest benefit.

Altogether, the CNRS spending authorizations will increase by 8.4 per cent in current francs against a predicted 3.6 per cent inflation. This is "a very useful and very important increase", says Papon, which will be distributed in line with the CNRS 7-year strategic plan prepared earlier this year (see *Nature* 315, 5; 1985). Life sciences "which have been a priority for 3 or 4 years" will see "a specific effort in general support for laboratory materials", but in recent months the franc has improved against the dollar, easing the problem in this area, said Papon. Engineering and social sciences will also receive special attention.

CNRS also plans to spend more on *valorization* — the process of making CNRS research profitable through links with industry, patenting and other measures — and to increase the budget of the international affairs division "especially for European cooperation". This may take the form of more exchange agreements, perhaps with Austria and Switzerland (where Papon is hoping to forge strong links), to complete agreements CNRS has now made "with every country in Western Europe".

There will also be more jobs at CNRS — another 300 for scientists and engineers in 1986. And Curien's ministry of research is aiming at the high target of 5 per cent annual scientific recruitment (as a proportion of present staff), to compensate for retirements and transfers. The immobility of French scientists, who rarely even stray from their home town, is still taxing the minister, however.

Other points in the budget include:

- Spending on the Agence Française pour la Maitrise de l'Energie (AFME), the French council for the support of initiatives in energy conservation and renewable resources, is to fall 3.7 per cent in current francs (some 7 per cent in real terms), reflecting a French energy glut due to nuclear power, and less pressure on national income due to lower oil prices.
- The medical research council INSERM will receive an 8.9 per cent increase (a little more than CNRS). The agricultural research council INRA is to get an 8.2 per cent increase.
- The budget of the space research council CNES, now developing the next, fully cryogenic version of the Ariane launcher, will be lifted another 23 per cent.
- Funds for developing French capability in electronics will also rise by nearly a quarter (23.8 per cent).
- The six biggest spenders in 1986 will be: CNRS (FF 8,951 million); Commissariat à l'Energie Atomique (CEA) (FF 7,044 million); CNES (FF 4,210 million); INRA (FF 2,233 million); and INSERM (FF 1,621 million).

Robert Walgate

Europe jostles for position in high-tech

Brussels

As part of a campaign to convince EEC foreign ministers of the important role that the European Community could play in promoting Europe's high technology, president of the European Commission, M. Jacques Delors, last week gave EEC foreign ministers their first look at the Community's new research and development "framework programme" for the years 1987 to 1991.

The new programme, intended to exploit the impetus given to technological cooperation by the French-inspired Eureka project, aims at striking a "new balance involving new topics". The areas covered include:

- The exploitation of space, mainly through remote sensing of resources and land-use in Europe and developing countries. This would involve setting up a user-network and some Earth stations, and would complement the facilities of the European Space Agency.
- The development of new energy-conserving and environmentally sound means of marine transport.
- The use of information and telecommunications technology to improve road transport in Europe.
- An as yet undefined project in aviation.
- The development of low cost educational tools for distance teaching (the Delta project).
- Improved research grants and social legislation to promote the mobility of research workers in the Community.

For most of these proposals, Delors hopes to have detailed plans worked out before the end of 1986.

As well as instituting new projects, the Commission is planning to speed up and extend existing EEC programmes. As a first step, the Esprit high-technology research programme should move into its next phase, involving demonstration and target-oriented projects. The recently relaunched medical research programme should include research on AIDS (acquired immune deficiency syndrome) and a cancer coordination project. And the biotechnology programme should be applying the new techniques to the problems of agriculture.

M. Delors sees room for, in the current EEC jargon, both "joint action" and "variable geometry". An example of the latter is the Eureka programme, which is basically a French initiative. The foreign ministers of the smaller EEC states (Netherlands, Belgium, Luxembourg and Ireland) have told M. Delors that they fear projects such as Eureka could leave them out in the cold, with the larger countries dividing the spoils between themselves. These smaller states are looking to the Commission to defend their interests and to become a focal point for future development of Eureka.

EEC research ministers will take up some of these points in an informal meeting on 23 October in the run up to the next Eureka interministerial conference, to be held in Hannover on 5 November.

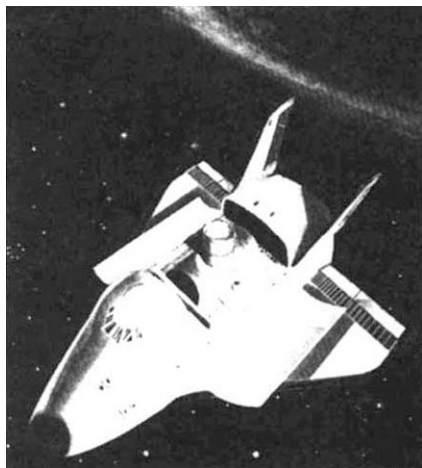
Anna Lubinska

Europe's shuttle

France forced to go it alone?

A DECISION which may upset the European partners of France is expected soon from the French government on who will be the prime contractor for the first European space shuttle. This "manned space plane" is Hermes, a miniature version of the US space shuttle, capable of carrying two to four astronauts into space and back again.

Hermes has been long in the planning by the French space agency CNES, and would fulfil a strategic need in Europe to have an independent, European means of placing and retrieving people in orbit, French space officials argue. Hermes would be launched by an up-graded, all-cryogenic version of the Ariane space launcher now under development at CNES (Ariane-5), and would have a payload varying from 1 ton and two men



Hermes — still only an artist's impression.

to 4.5 tons and four men according to orbit. Unlike the US space shuttle, Hermes will be designed to go into polar orbit at relatively high altitudes (800 km) as well as in low equatorial orbit; the fuel costs of reaching high polar orbit would minimize Hermes' payload. The shuttle should carry 4.5 tons and 4 people around the Equator — but only 1 ton and two people around the poles.

Full-scale development costs have been estimated at some FF 1,400 million (around £ 140 million). Frederic d'Allest CNES director-general (and president of Arianespace, Ariane's commercial wing), has been touring Europe attempting to raise enthusiasm and cash for the project, but the French government now seems so committed to the project that France may continue alone if full backing is not forthcoming. Last January, the European Space Agency in its first full ministerial meeting for years "noted with interest the French decision to undertake the Hermes programme" and asked to be "kept informed" of progress. And if all goes well, the European Space Agency (ESA) may adopt Hermes as one of its "optional" commercial programmes (like Ariane)

sometime next spring, at about the time when the agency also has to make a decision on full participation in the US space station project (now in the detailed definition phase).

Meanwhile, France is pushing ahead fast in an attempt to define the French commercial structure behind Hermes. At present the battle is fittingly between aircraft manufacturer — Dassault — and Aérospatiale, which assembles Ariane and the French nuclear missiles the SSBS and MSBS. Dassault has the expertise in flying; Aérospatiale in re-entry technology (for nuclear warheads), rocket assembly and space data systems. Either could take the prime contractorship for Hermes and both are competing for it; and both, in principle, should be involved in its construction.

However, there appears to be some delay in announcing the choice of prime contractor — and this may not be unconnected with the way Dassault just a few weeks ago scotched French participation in the next-generation European fighter-

plane project, the EFA. After years of negotiation, France dropped out of the collaboration — which will now take place between Britain, West Germany, Italy and Spain — arguably because Dassault sought the lion's share of the work. (Another reason was that Dassault also had an eye on the export market, and wanted a lighter plane than would have suited the North Atlantic Treaty Organization's European theatre, for which the EFA was primarily destined.) But whatever the cause, Dassault effectively wrecked its European credentials over the EFA. There is therefore some doubt among European partners about participation with the company on another, even more adventurous, European project.

As with the EFA, France seems likely to make her own, and independent, choice irrespective of outside pressures. As Roger Lesgards, president of the Société Européenne de Propulsion (SEP) that constructs Ariane's motors, wrote recently: "It is evidently indispensable in the European space effort that France should take the first rank". If that rank means Dassault, then the feeling in Paris seems likely to be: so be it. **Robert Walgate**

Genetic engineering

Restrictions on industry eased

Hamburg

THE safety regulations for gene technology, which were adopted in West Germany in 1978, are now being revised, with the intention of relaxing the tight restrictions which are hampering companies trying to scale-up biotechnology to the industrial level. The new regulations should be published early in 1986. The Federal Minister for Research, Dr Heinz Riesenhuber, hopes that these relaxations will help the German biotechnology industry to become more competitive internationally. In June of this year, he promised the industry DM 1,000 million as support for the years 1985 to 1989 (*Nature* **316**, 287; 1985).

Until the new regulations become effective, West German companies are allowed to conduct large-scale fermentations of transformed organisms (in vessels larger than 10 litres) only after obtaining special permission from the *Bundesgesundheitsamt* (Federal Health Office). The new guidelines will define general criteria and conditions under which large-scale fermentations are to be permitted, and thus prevent the unfairly harsh treatment of companies eager to move on to commercial production. The use of gene-manipulated organisms in field research will still only be allowed after the issue of a special licence, however.

The 1978 regulations applied only to research projects which were financially supported by the Federal Research Ministry, and this will remain the case for the

new regulations. Riesenhuber, however, expects that all industrial concerns will adopt the new rules, and indeed, he has already received promises of cooperation from industrial representatives. In his opinion the revisions are needed urgently because two applications for industrial production capacities (Hoechst for insulin, and Bioveron for interferon) have already been submitted to the Federal Health Office.

The new regulations have met with some opposition, however, and Riesenhuber has been criticized for not waiting for the report of the Enquête Commission on gene technology, established by the *Bundestag* in autumn 1984. The social-democratic chairman of this commission, Wolf-Michael Catenhusen, agrees that there is a need for new regulations, but says that the investigation of his commission on safety risks should be given adequate consideration. A first hearing took place at the end of September, a second hearing on questions of safety in the licensing of genetically modified organisms will take place on 21 October, but the Commission will decide the matter only at the end of 1986.

The Ministry for Research has stressed that the revised safety recommendations are not intended to subvert those from the Enquete Commission. Similar, but internationally applicable recommendations are being formulated by the Organization for Economic Cooperation and Development. **Jürgen Neffe**

Velikovsky's theory

SIR—In his review of Henry Bauer's *Beyond Velikovsky* (*Nature* 25 April, pp. 692–693), Owen Gingerich claims that Peter Huber has “provided evidence from Babylonian records that Venus in 1600 BC behaved just as it did in the centuries immediately before our own era” — and therefore essentially as it does today. What Gingerich is trying to show is that Velikovsky's theory that Earth and other planets experienced orbital changes as recently as twenty-seven centuries ago must be false. The Venus or Ninsiana tablets to which Huber and Gingerich are referring record invisibilities of Venus at inferior and superior conjunctions over a number of years. But the lengths and spacings of those invisibilities are grossly incompatible with what is seen today. This fact was discussed by Velikovsky in 1950 (*Worlds in Collision*, pp. 198–199), as evidence supporting his theories.

As far as I know, no serious investigator of Velikovsky's theories has ever “worked hard to discredit the Babylonian data”, as Gingerich claims. On the contrary, Raymond C. Vaughan and I, in a series of publications going back to the early 1970s, have worked to vindicate the Babylonian data. Those who want to discredit the data are precisely people like Gingerich and Huber, who have to reject about two-thirds of the data. (See *Kronos* X:2, pp. 1–12.) When these records are taken seriously and are analysed in terms of the orbits that they imply, one of the findings is that Earth was on an orbit whose eccentricity was at least five or six times the present value. Thus the Babylonian observations of Venus or Ninsiana strongly support Velikovsky's theory that Earth has undergone catastrophic orbital change within historical times.

LYNN E. ROSE

State University of New York,
Department of Philosophy,
Baldy Hall,
Buffalo, New York 14260, USA

Price of star wars

SIR—The primary aim of President Reagan's Strategic Defense Initiative is to attack Soviet Intercontinental Ballistic Missiles with particle-beam (or other futuristic) weapons during the boost phase. At this stage, missiles from, for instance, the major Soviet launch site at Plesetsk are likely to be near the Finnish border. A successful strike will either damage the guidance electronics or detonate the high-explosive trigger of the nuclear warhead (without necessarily causing a nuclear explosion). Either action could direct radioactive material onto Europe; the explosive action would also scatter radioactive elements into the atmosphere.

A Soviet first strike might involve some 5,000 warheads. If only 20 per cent of

these, each containing about 10 kg of plutonium-239 (half-life 2.4×10^4 yr), disintegrate (without a nuclear explosion) in the Northern Hemisphere, about 10^{13} lethal doses (if inhaled or ingested) of radioactive material will be released — about 5,000 per person in the Northern Hemisphere. Is this not a high price to pay for the President's star wars dream?

R. V. HARROWELL

Devana, Cardinal's Green,
Horseheath, Cambridge CB1 6QX, UK

Sahel famine

SIR—As a layman, I suggest that the plan by A. R. E. Sinclair and J. M. Fryxell (see *Nature* 5 September, p. 13) for the recovery of Sahel fertility omits two essential items. First is the need for political stability and second for social approval of the proposals. In a previous major incident, the treatment of the dustbowl in the southern plains of the United States in the 1930s, the political stability was there but social acceptance was not, and recovery was organized by agronomists and not ecologists.

Man is not forced to accept traditional solutions to his problems; his nature enables him to invent new ones.

E. NICHOLL

Llangwmville,
Llanrothal,
Monmouth NP5 3QL, UK

Japan's performance

SIR—In their illuminating Commentary (*Nature* 316, 587; 1985) based on a comparative study of the scientific performance of major industrialized countries, Irvine, Martin and Turner repeatedly emphasize the remarkable achievements of Japan in terms of both publication and citation over the period 1973–82. I wonder whether their praise is justified. They give only the percentage increases of the share in publication and citation (40 per cent and 65 per cent respectively) and apparently base their conclusion on these numbers. The ratio of two numbers can become large for two distinct reasons; either the numerator is large or the denominator is small. The percentage increase can look unduly large for some non-intrinsic reason, for example Japan-based literature is less comprehensively surveyed in 1973 than in 1982, or the share itself is so small that its small increase leads to a large percentage increase. For the purpose of learning whether such factors have to be remembered in accepting the conclusion, I request the authors to publish the share of individual countries either for the year 1973 or 1982 or for both.

KATUHISA SUZUKI

Department of Physics,
Osaka University,
Toyonaka 560, Japan

Over-exposure?

SIR—*Nature* has gone too far this time—I refer to the article on volcanism in the South Pacific (*Nature* 316, 507–511; 1985), with 19 authors from 18 institutes. Altogether the title of the article and names and addresses of the authors occupy 50 per cent of a page, over 10 per cent of the total article. For what purpose? A cynical reviewer might suggest the intention was that the maximum number of scientists, and for that matter institutions, should get referable exposure in *Nature*. *Nature* column inches are extremely valuable and are surely designed for communicating science. Are they being subverted for self-promotion?

JOHN PALFREYMAN

Department of Molecular
and Life Sciences,
Dundee College of Technology,
Bell Street, Dundee DD1 1HG, UK

PhD theses

SIR—I am not sure whether Beverley Halstead's statement (*Nature* 29 August, p. 760) that “if the material is not suitable for publication, then it was not worth doing in the first place” would find favour with those who spend many years painstakingly developing a new area before it can yield publishable results. This should not, however, be a task for a PhD student, and so the requirement suggested by Dr Halstead for academic publication before the reward of a PhD seems reasonable.

But were this to be insisted on, some measures would have to be taken to ensure that students were given projects that stood a reasonable chance of success given sufficient talent and enthusiasm on the part of the student. The departments of our universities are full of postgraduates working under inadequate supervision on projects that have little hope of success. Supervisors too often seem to regard the arrival of a student as an opportunity to move into a new field with which they themselves are not familiar and to which they would not devote their own efforts. The end result is frequently a poor thesis, cobbled together from bits and pieces, or, worse, a failed and disillusioned student.

A few years ago, the Science and Engineering Research Council published a small booklet called *A Guide to Good Supervisory Practice*. Its contents were aimed at both supervisor and student. It contained a number of suggestions which, if thoughtfully applied, would help to ease the labour pains involved in the production of that troublesome yet coveted child, the PhD thesis. It is sad that, all too often, one or both of the involved parties have less interest in the birth than might have been expected given the difficulties in the gestation.

P. A. LUND

Department of Biochemistry,
University of Bristol, Medical School,
University Walk, Bristol BS8 1TD, UK

Cold water on icosahedral symmetry

Linus Pauling has produced an alternative explanation of the observation that solid manganese-aluminium alloy may have 5-fold symmetry on the atomic scale. How can the two views be reconciled?

DR Linus Pauling, the long-standing subversive, has once more thrown a cat among some pigeons. His article on page 512 of this issue is several things, but not least a reminder that his career began in the 1920s as that of a crystallographer. The trouble that Pauling's article will now stir up is simply that it denies the need for speculation about "a new kind of matter" to explain observations of 5-fold rotational symmetry in the structure of solid specimens of a manganese-aluminium alloy.

The first observations were reported at the beginning of last year by Schechtman *et al.* (for this and other references, see Pauling's article). The alloy, which in composition approximates MnAl_6 , is made in small quantities by fast-quenching molten metal. Most simply, X-ray diffraction in the appropriate directions yields patterns of spots with clear 5-fold symmetry. The observations have been repeatedly confirmed and ingeniously explained by the notion of quasi-crystalline matter. For all practical purposes, Pauling now says that these schemes are unnecessary.

The elementary textbooks carefully explain why the only kinds of rotational symmetry consistent with the translational symmetry of a crystal lattice are those with 2-, 3-, 4- and 6-fold axes. Five-fold symmetry axes are ruled out because if rotation by a fifth of a full circle will leave the crystal lattice unchanged, rotation by two-fifths of a full circle will not. But Levine and others have sought to construct three-dimensional atomic frameworks for solid materials by a technique recognized, after the event, as a generalization of the Penrose covering of the two-dimensional plane with tiles shaped like one or other of two distinct rhombi (see *Nature* **313**, 263). The Penrose tiling has some strict 5-fold axes of symmetry (of rotation perpendicular to the plane). Moreover, all edges marking the boundaries between tiles have orientations belonging to a finite set of directions in the plane. The price paid for pentagonal symmetry is that strict translational order is lost.

These are hallmarks of the 'quasi-crystal', as elaborated during the past year. A three-dimensional framework of atoms can be built so as to have true 5-fold symmetry, because in Penrose tiling neighbour-neighbour bonds have fixed orientation but translational periodicity is replaced by quasi-periodicity. All lattice points are specified as integral combinations of vectors with lengths incommensu-

rate with each other. It is generally agreed that the theories are quite fun.

Pauling's approach is classical and sceptical. Why not, instead, see what traditional model building might accomplish? Quite apart from the significance of the result, the argument is an illustration of how intuition, experience and apparently universal knowledge can be used to fashion a complicated and unknown crystal structure out of thin air.

The interdiction against 5-fold symmetry in crystal lattices does not of course imply that atomic arrangements, molecules perhaps, with 5-fold symmetry, cannot form solids but merely that the symmetry of the groupings cannot be imparted to that of the structure as a whole. So Pauling starts with simple structures which do have 5-fold symmetry, the known icosahedral intermetallic complexes such as that of tungsten and aluminium. The 5-fold symmetry of a regular icosahedron is self-evident in that the twelve faces of this regular polyhedron are regular pentagons, but apart from the 5-fold axes through each face, there are also 3-fold axes through each vertex.

The conceptual model-building from this point on depends to some extent on the happy chance that the tetrahedral angle (109.47°) is not very different from the included angle of a regular pentagon. To give the material a chemical composition corresponding to the Schechtman alloy, Pauling assumes each elementary icosahedron shares four faces with others. By bending the angles a little, he shows that five icosahedra can make a ring, or that twenty can yield a structure with the symmetry of a dodecahedron.

For the Schechtman alloy, what matters most, in Pauling's argument, is that the spots on the diffraction diagrams correspond to those the model predicts. Remarkably, the predicted positions are those determined by the simplest guess about the radii of the cores of manganese and aluminium. That the intensities of the spots should also vary systematically as predicted by the model is a clinching piece of circumstantial evidence.

So is the Schechtman alloy a regular crystal in the Pauling mould, or a quasi-crystal as described by Levine and others? There is no doubt that the observations are unshakable evidence of 5-fold symmetry and that the symmetry extends over regions large enough for the diffraction spots to be sharp, which leads to the estimate that the microcrystals responsible

have dimensions measured in micrometres. Pauling, on the other hand, has shown how it would be natural for a manganese-aluminium alloy with the Schechtman composition to form into arrangements which, on the scale of some tens of ångströms, resemble that of a dodecahedron. On Pauling's model, this structure would form a regular crystal with a cubic unit cell of dimensions $26 \text{ \AA}$. One way of visualizing the result is as a series of concentric shells built up of icosahedra sharing common faces which, because of their inherent trigonal symmetry, are arranged in space relative to each other much as are the carbon atoms in diamond. But then, the argument goes, because of the strain entailed in bending the tetrahedra into pentagonal angles and, more important, the twisting involved in forming chains of icosahedra into close rings, the time comes when it is energetically advantageous to start building a new crystal on each of the twenty faces of the underlying dodecahedron. The result, Pauling says, should be a speck of solid material built from a seed with dimensions of the order of $100 \text{ \AA}$ supporting up to 20 crystal twins, each with a different orientation. Given the size of the underlying structure, it may not be unreasonable to suppose these structures growing to micrometre dimensions.

So how are the two structures to be reconciled? Pauling has one practical suggestion: more accurate X-ray diffraction measurements with monochromatic X rays so as to be able to index a much larger number of diffraction spots and then more accurately to compare prediction and observation. If such experiments confirm his ingenious model, it will seem a triumph for the intuitive crystal model-building of the traditional kind.

If the Schechtman alloy is not, after all, an example of a quasi-crystal structure in the sense introduced by Levine, where should people look for realizations of structures of that novel kind, which are not inherently implausible? And what relationship is there between the atomic framework that Pauling has built on paper and that suggested by the advocates of quasi-crystals? It would be a gigantic task to assemble all the atomic coordinates of a crystal in the Pauling mould; might it even be that, if there were done, it would turn out that the two descriptions are but different ways of describing the same kind of arrangement? That would save a lot of trouble.

John Maddox

Phase transitions

Cosmology in the laboratory

from T. W. B. Kibble

AT FIRST sight it may be surprising that useful information about the processes leading to the formation of the largest structures in the Universe could come from small-scale laboratory experiments. Yet just such a bold proposal is made by W. H. Zurek on page 505 of this issue¹.

The proposed experiment is a very simple one. Apply a rapid pressure quench to a sample of liquid helium in an annular container to take it quickly through the transition from normal to superfluid. Zurek predicts that the sample will acquire a random non-zero circulation velocity which should be readily measurable. If confirmed, this will be of considerable interest in its own right in terms of the physics of superfluids, but the motivation for the proposal comes from the need to test our ideas about cosmic strings — a remarkable example of fruitful fertilization between very diverse disciplines.

Phase transitions occurring in the very early stages of the history of the Universe (only a fraction of a second after its birth) may generate various topologically stable defects² such as monopoles, domain walls and in particular strings, which are analogous to the quantized vortex filaments in superfluid helium or the flux tubes in superconductors. It was suggested by Zel'dovich that these cosmic strings might provide the initial density perturbations that lead to gravitational condensation of galaxies. This idea was refined by Vilenkin^{4,5} and others into a very attractive theory of galaxy formation. Recent, as yet unpublished, work by Turok has shown that it can explain rather well the correlations of large clusters. The theory depends, however, on an assumption about the density of strings produced at the phase transitions^{6,7}, which is of course quite inaccessible to direct observations. Zurek's idea is to test the assumption on superfluid helium.

When liquid helium undergoes the transition from normal to superfluid, the complex order parameter ψ , which can be thought of as the wave function of the Bose condensate, acquires a non-zero value. If the transition is sufficiently rapid, the phase of ψ will vary randomly from one region to another and vortex lines will be produced. These are trapped regions around which the phase varies by a multiple of 2π . Since the superfluid velocity is proportional to the gradient of the phase, the lines carry quantized vorticity.

The vortex lines once produced will evolve rapidly and leave no direct evidence of their initial distribution. But Zurek points out that if an annular geometry is chosen, there may be a non-zero net circulation around the annulus.

Because vorticity is quantized, this flow will persist and may lead to a circulation velocity in the superfluid of the order of millimetres per second, which should be readily measurable.

Even if this prediction is confirmed in relation to superfluid helium, why should we believe that it tells us anything about cosmic strings? After all, the transition temperatures differ by factor of perhaps 10^{28} . The answer lies in the universality of behaviour near a phase transition. Liquid helium is well described qualitatively by the Landau-Ginsberg model. The relativistic version of this is the Higgs model, the paradigm for discussions of spontaneous symmetry breaking in field theories. Provided the symmetries of the problem are the same, we expect the same qualitative behaviour, despite the very different scales. Vortex lines should be a good analogue of at least one class of strings.

With regard to the most popular type of cosmic string, however, there are impor-

tant differences. Most significant is the fact that the symmetry involved in the superfluid case is a global one, not a local gauge symmetry. In particular, there is no gauge field coupled to the relevant current. In this respect, a better analogue would be a flux tube in a superconductor. It would be fascinating if one could perform the Zurek experiment in a superconducting ring, but it might be very hard to induce a sufficiently rapid transition. In helium, the existence of another thermodynamic variable, the pressure, is crucial. Magnetic field could not be used in a similar way, because the goal would be to generate magnetic field spontaneously.

Particle physicists and cosmologists will eagerly await the results of the Zurek experiment. The expected positive result would be a considerable boost for the string theory of galaxy formation. □

1. Zurek, W. H. *Nature* **317**, 505 (1985).
2. Kibble, T. W. B. *J. Phys. A* **9**, 1387 (1976).
3. Zel'dovich, Y. B. *Mon. Not. R. astr. Soc.* **192**, 663 (1980).
4. Vilenkin, A. *Phys. Rev. Lett.* **46**, 1169 (1981); **46**, 1496 (1981).
5. Vilenkin, A. *Phys. Rev.* **24**, 2082 (1981).
6. Kibble, T. W. B. *Nucl. Phys.* **252**, 227 (1985).
7. Albrecht, A. & Turok, N. *Phys. Rev. Lett.* **54**, 1868 (1985).

T. W. B. Kibble is Professor of Theoretical Physics, Imperial College, Prince Consort Road, London SW7 2BZ, UK.

Human genes

Mapping hereditary disorders

from Albert de la Chapelle

THE rate of progress with which human genes are being allocated to identified regions of individual chromosomes is rapid. In four years, the number of mapped genes has increased nearly three-fold to over 900. Sixteen mapped genes had been cloned four years ago; the present figure is 249. And the number of regionally localized DNA segments ('probes') that are not known to represent genes but are an important aid to mapping has risen from some 50 to over 800. These figures were the raw facts of a recent workshop* that illustrated how knowledge of the human gene map helps to clarify the molecular basis and pathogenesis of many mendelian disorders. Moreover, with the aid of the gene map, disorders can be diagnosed before they are molecularly understood.

Cystic fibrosis (CF), one of the most common recessively inherited mendelian disorders of Western populations, has an incidence at birth of approximately 1 in 2000 (with a remarkable gene carrier frequency of about 1 in 20). Nevertheless, ignorance about the CF gene has been virtually total. The first lead, reported at Helsinki, is evidence of a linkage between the CF gene and the *PON* gene for the enzyme paroxonase¹. The measure of the

reliability of the linkage, the 'lod score', is 3.38, where lod scores in excess of 3.0 are considered 'formal proof' of linkage. The data suggest that the genes for *CF* and *PON* are located approximately 10 recombination units apart on the same, so far unidentified, chromosome. Once the *PON* gene has been cloned, it should be possible to determine the chromosome and band where it and *CF* are located by the methods of classical somatic-cell genetics or *in situ* hybridization. Then chromosome-specific libraries will be screened for closely linked probes, which should lead more quickly to the *CF* gene than chromosome 'walking' or 'hopping' from *PON* to *CF*, since the genetic distance of 10 recombination units is long, corresponding to some 10^7 base pairs.

If the *CF-PON* linkage is confirmed, it will certainly represent a breakthrough in CF research but, unfortunately, the *PON* phenotypic polymorphism is quantitative rather than qualitative. It remains to be seen how much overlap there is in the observed bimodal distribution of paroxonase concentrations in the plasma and whether it is really due to alleles at the structural *PON* locus.

Another inherited disease of unclear aetiology and pathogenesis is adult polycystic kidney disease (PKD1). Its gene frequency is approximately 1 in 1000 and, as

*The Eighth International Human Gene Mapping Workshop, Helsinki, Finland, 4-10 August, 1985. Proceedings will be published as vol. 40 of *Cytogen. Cell Gen.* this year.

it is dominantly inherited, the incidence of the disease equals the gene frequency. A close linkage between *PKD1* and the α -globin gene was reported in Helsinki² and is published on page 542 of this issue³. Thus *PKD1* must reside on the short arm of chromosome 16. The recombination distance is 5 units and the lod score is a very convincing 25.85. Very close linkage to the gene for phosphoglycolate phosphatase is also found.

These findings mean that genetic counselling in affected families should soon be possible; when probes bridging *PKD1* are identified, highly accurate prediction will be possible. Thus, as in a number of other mendelian disorders, including the haemoglobinopathies, phenylketonuria and Duchenne's muscular dystrophy, diagnostic procedures may become available before the molecular biology is fully clarified. One reason why high lod scores are obtained in *PKD1* is the existence of a cloned DNA segment containing a highly polymorphic region close to the 3' end of the α -globin gene cluster³. Highly polymorphic DNA probes (with polymorphism information content, or PIC, values approaching 1.0) are so useful in gene mapping that only a few are needed to map the whole genome. What "few" signifies is disputed, but for a comparatively brief stretch of DNA, such as the short arm of chromosome 16, five such probes, if evenly placed, would include at least one that was closely linked to *PKD1*. Presently, 13 DNA probes are mapped to chromosome 16, which is few compared with the 69 probes for chromosome 21, 49 for chromosome 4 and 214 for the X chromosome.

Since many of the genes now mapped are related to human disease, diagnostic procedures can be rapidly developed. Examples where this should now be possible include von Willebrand's syndrome, whose gene is on chromosome 12 (refs 4 and 5) and Wilson's disease, whose gene is mapped to chromosome 13 (refs 6 and 7).

Several genes on chromosome 19 may serve the same purpose and extensive linkage data for that chromosome were presented by many groups. A likely gene order that emerged is: familial hypercholesterolaemia, Lewis blood groups, complement C3, Landsteiner-Wiener blood group, peptidase D, myotonic dystrophy, secretor, apolipoproteins C and E, Lutheran blood group. Twenty-two DNA clones, 10 of which are polymorphic, are available on chromosome 19. Thus, several genes related to major disorders and lying within a reasonably short stretch of DNA are much more accessible than they were.

The proceedings of the Helsinki workshop, which will be published before the end of the year as vol. 40 of *Cytogenetics and Cell Genetics* contain listings of all 808 DNA clones (249 genes and 559 probes) so far regionally mapped. The listings contain not only a brief description of each

probe, but also information regarding availability and source. There is also a critical comprehensive listing of breakpoints involved in recurring chromosome aberrations associated with specific neoplasms. Eighty-five such recurring breakpoints in 31 different neoplasms are now known, compared with 44 breakpoints in 17 neoplasms two years ago. These breakpoints are important not only for understanding cancer but also because they represent a new tool for gene mapping. □

1. Eiberg, H. *et al.* *Human Gene Mapping 8* Abstr. 100 *Cytogen. Cell Gen.* (in the press).
2. Reeders, S.T. *et al.* *Human Gene Mapping 8* Abstr. 450 *Cytogen. Cell Gen.* (in the press).
3. Reeders, S.T. *et al.* *Nature* **316**, 542 (1985).
4. Verweij, C.L. *et al.* *Human Gene Mapping 8* Abstr. 374 *Cytogen. Cell Gen.* (in the press).
5. Ginsburg, D. *et al.* *Science* **228**, 1401 (1985).
6. Frydman, M. *et al.* *Proc. natn. Acad. Sci.* **82**, 1819 (1985).
7. Cox, D.W., Weitkamp, L.R., Lewis, M. & Sass-Kortsak, A. *Human Gene Mapping 8* Abstr. 76 *Cytogen. Cell Gen.* (in the press).

Albert de la Chapelle is Professor of Medical Genetics, University of Helsinki, Haartmaninkatu 3, 00290 Helsinki 29, Finland.

Meteorites

Evidence of Martian origins

from R. O. Pepin

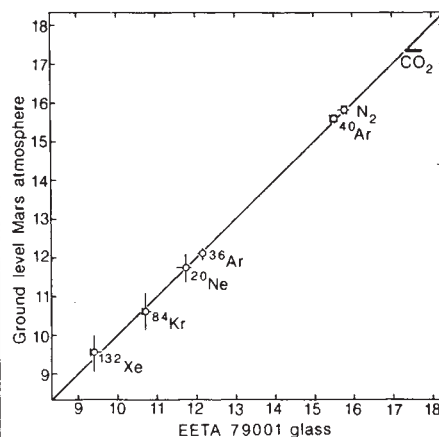
GONE is the old orthodoxy that meteorites come only from small bodies in the Solar System — asteroids and perhaps comets. Most do, but not all. Among the thousands of meteorite fragments disintegrated from Antarctic ice flows are three pieces of the Moon, instantly recognized for what they are by their unambiguous kinship to the Apollo lunar samples. And now another Antarctic specimen has opened up even more remarkable perspectives on the possible provenances of meteorites. Its name, EETA 79001, may be prosaic but its chemical and isotopic signatures are not: they firmly link this meteorite, through data returned to Earth from the Viking spacecraft, to an origin from the surface of Mars. On page 509 of this issue, Ott and Begemann ask if the same can be said of its fellows¹.

EETA 79001 is a shergottite, one of a group of eight puzzling stones collectively called the SNC meteorites (after Shergotty, Nakhla and Chassigny, the type specimens for the three subgroups, see ref. 2). They seem, on petrological, chemical and isotopic grounds, to be more closely related to each other than to other classes of stony meteorites; they are unusually young — for meteorites that is — their chronology is complicated, but they have probably crystallized from melts within the past 1,500 Myr; and they are most readily understood as products of differentiation within a large, thermally active planet. In fact, in many of their general characteristics these meteorites look like terrestrial basalts.

In 1979, Sherlock Holmes-like logic led Wasson and Wetherill³ to conclude that Mars was the least improbable parent body for the SNC meteorites, a view supported by Wood and Ashwal⁴ on entirely inferential arguments. The breakthrough that transformed the basis of the entire debate from circumstantial arguments to direct and quantitative evidence was Bogard and Johnson's discovery^{5,6} that glassy nodules imbedded in the EETA 79001 rock contain a unique noble gas component — unique in the sense that its elemental and isotopic composition of

neon, argon, krypton and xenon does not resemble that trapped in other meteorites. But it does resemble, in striking detail, the compositional pattern measured by the Viking spacecraft for these gases in the martian atmosphere.

Bogard and Johnson's report at once suggested a second experiment, namely, to search the EETA 79001 glass for the isotopically heavy nitrogen that distinguishes the atmosphere of Mars from virtually all other volatile reservoirs in the Solar System. Such analysis⁷ did in fact reveal very large enrichments of ¹⁵N in the



Comparison of the relative abundances of various gases trapped in EETA 79001 glass with samples of the martian atmosphere taken by the Viking spacecraft. The number of particles per cubic centimetre is shown on a log — log plot.

trapped gases. Subsequent work has sharpened the precision of the noble gas measurements⁸ and extended the data base to include carbon dioxide^{9,10}.

The figure shows a comparison between current information on volume abundances of all these gases in the EETA 79001 glass and Viking's measurement¹¹ of ground-level atmospheric abundances on Mars. Within their respective uncertainties, both the relative and absolute amounts of noble gases, nitrogen and carbon dioxide in the two data sets can be seen as identical. In my view, this is *prima*

facie evidence for the origin of EETA 79001 on or near the surface of Mars.

The next question is exactly that posed by Ott and Begemann¹: is there direct evidence in the noble gases that ties the SNC meteorites together and thus to the same parent body? If EETA 79001 is from Mars, are they all? As pointed out earlier, these meteorites do resemble each other in many respects. One of the most powerful ways to explore such interrelationships is by use of isotope ratios; perhaps the most convincing evidence at present for a common SNC parent is their shared distinctive oxygen isotope signature¹².

In their study, Ott and Begemann look for the isotopic and elemental signature of the EETA 79001 noble gas component in the Shergotty, Nakhla and Chassigny meteorites. The results are mixed — literally. While not seen in pure form (undoubtedly due, as noted by the authors, to the absence of the unusual macroscopic glass nodules that are the principal trapping sites in EETA 79001), there is convincing evidence for the presence of these

noble gases in Shergotty, where they appear mixed with a second kind of gas component that may be represented, more or less, by that in the Chassigny meteorite.

Wiens *et al.*¹³ have found just this situation within EETA 79001 itself, with clear indications of two independent nitrogen and noble gas components. One, dominant in the glass, is attributed to the martian atmosphere; the other seems likely to derive from the mantle source region for these basalts. The only real departure from a coherent two-component mixing pattern in Ott and Begemann's data appears in Fig. 1b of their paper, where the points for the Nakhla meteorite fall above the mixing line. This could require a third gas component, rich in ¹²⁹Xe, as the authors suggest. Another possibility is that the high ¹²⁹Xe/¹³²Xe ratio is due, as in Shergotty, to the presence of atmospheric gases, but with relative elemental abundances severely fractionated by whatever the processes were that led to their incorporation into Nakhla. Fractionation of the Kr/Xe ratio by about a factor of five would

be needed to account for the data. It is a pity that the large amount of spallation-produced argon in Nakhla masks so completely the trapped ³⁶Ar abundance. If this were not so, the prediction of a very low trapped Ar/Xe ratio resulting from mass fractionation of this order could be tested.

In large part, the scientific excitement generated by the SNC–Mars association comes from the potential for using these meteorites to address questions about Mars that cannot be answered, or even adequately explored, using only the data now in hand from Earth-based and Viking observations. One important class of problems concerns the initial inventory and subsequent history of volatile elements and compounds — the noble gases, nitrogen, carbon dioxide and water — on Mars. Here, for believers, the SNC meteorites will certainly play a leading role in meeting the challenge of understanding the origin and evolution of the martian atmosphere, and the relation of martian volatiles to those on Earth, Venus and the more usual meteorite parent bodies. One example is Ott and Begemann's discussion of a central question: how, if at all, do noble gases of the type found in Chassigny relate to the present atmosphere of Mars? Did the atmosphere derive from this kind of mantle reservoir or are we dealing with some independent source of atmospheric gases? The fundamental difficulties they encounter in deriving the current atmosphere from Chassigny-type sources may force attention to an alternative model: a gas-bearing veneer of primordial, late-accreting matter that could well be unrepresented in the contemporary meteorite population.

All this should not be taken to mean that there are no outstanding problems with a martian origin for the SNC meteorites. The gravitational potential well of Mars is deep, and calculation of the mechanics of a meteorite impact on the surface reveals difficulties in ejecting from the planet fragments of the crust a metre or more across¹⁴. The residence times of these meteorites in space, given by their cosmic-ray exposure ages, imply ejection at different times and therefore by more than one impact event. This is troublesome if the chemical and chronological similarities within the SNC suite are interpreted to mean that they all derived from a very restricted area of the martian surface, so that more than one large meteorite strike into the area during the past 10 Myr or so becomes improbable.

Discovery of meteorites from the Moon has verified that at least fragments can be liberated from relatively large bodies without being destroyed in the process. Given eight SNC meteorites from Mars, however, several of which have been observed to fall in the past two centuries, there remains the question of whether we should not be up to our necks in lunar meteorites — that is, what would be the expected relative fluxes of objects from

Paul J. Flory (1910–1985)

ON 9 September, Paul Flory died of a heart attack at his holiday home in Big Sur, California three months after his many friends, colleagues and past students had celebrated his 75th birthday with a scientific symposium at Stanford University.

Paul Flory was awarded the Nobel Prize in chemistry in 1974, in recognition of enduring contributions to polymer science, a field that he, as much as anyone, helped to found. To salute his achievements, Stanford University Press this year published three massive volumes entitled *Selected Works of Paul J. Flory*, edited by past and present associates. A notion of the impact of Flory's contribution on the development of polymer and related sciences can be gained by turning the pages of this magnificent testimony to human endeavour, spanning half a century of unceasing achievement.

Flory published his first papers on the kinetics and mechanism of polymerization and molecular weight distributions in high polymer systems in 1936, while employed in the group of Wallace H. Carothers, the inventor of nylon, in the research laboratories of the Du Pont Company. At that time polymer science had just come into being, following the efforts of Hermann Staudinger to prove the existence of large molecules to a reluctant chemical establishment. With uncanny intuition Flory enunciated the rules governing the generation of high polymers and those describing their behaviour in solution and in the solid state. It is thanks to his insights and those of his illustrious contemporaries, in particular Werner Kuhn, Peter Debye and John Kirkwood, that we have come to a quantitative understanding of macro-

molecular behaviour. Flory's work and achievements cover such diverse aspects as solution thermodynamics of polymers and mixtures, hydrodynamics, the rules governing melt viscosities and the transition to the glassy state, crystallization of polymers from melts, chain configuration and conformational analysis, rubberlike elasticity, semicrystalline rigid polymers and liquid crystals. To Paul Flory we owe an understanding of the excluded volume in solutions of macromolecules and the concept of the theta, or Flory, temperature, at which the excluded volume vanishes and non-idealities in polymer solutions are internally compensated. At the Flory temperature, macromolecular chain configurations are unperturbed in solution. Many years before neutron scattering from deuterated macromolecules made it possible to verify his prediction, Flory had postulated that macromolecules in the amorphous bulk state would also assume unperturbed configurations. Flory's book *Principles of Polymer Chemistry* published in 1953, remains the classical reference in polymer science; to it was added in 1969 *Statistical Mechanics of Chain Molecules*.

Paul Flory's creative activities continued up to the time of his death. He was a "severe" teacher, an outstanding personality and a man of high principles. He had strong convictions concerning the responsibilities of scientists in the service of society and his extramural activities were manifold. I would like in particular to stress his efforts on behalf of dissident scientists in Eastern Europe, efforts which he pursued with great vigour and devotion. Personalities of his intellectual and moral calibre are indeed few. Henryk Eisenberg

the Moon and Mars and why have we seen so few from the Moon? All these unanswered questions, and others, stand as challenges to the SNC-Mars connection. But there is a lot we do not know in most of these areas, and the geochemical link to Mars, at least for EETA 79001, is very strong indeed.

The final proof lies on the surface of Mars, awaiting collection — with luck from a shergottite outcrop — by a future generation of astronauts. In the meantime the SNC meteorites will be thoroughly dismembered in laboratory explorations, at a pace that is certain to be accelerated by contributions as stimulating as that from Ott and Begemann. Clearly, we shall learn much about the SNC parent body. □

1. Ott, U. & Begemann, F. *Nature* **317**, 509 (1985).
2. Wright, I.P. & Grady, M.M. *Nature News and Views* **315**, 367 (1985).
3. Wasson, J.T. & Wetherill, G.W. in *Asteroids* (ed. Gehrels, T.) 926 (Univ. Arizona Press, Tucson, 1979).
4. Wood, C.A. & Ashwal, L.D. *Proc. 12th. lunar planet. Sci. Conf.*, 1359 (1981).
5. Bogard, D.D. *Meteoritics* **17**, 185 (1982).
6. Bogard, D.D. & Johnson, P. *Science* **221**, 651 (1983).
7. Becker, R.H. & Pepin, R.O. *Earth planet. Sci. Lett.* **69**, 225 (1984).
8. Swindle, T.D. et al. *Geochim. cosmochim. Acta* (in the press).
9. Carr, R.H. et al. *Nature* **314**, 248 (1985).
10. Wright, I.P. et al. *Geochim. cosmochim. Acta* (in the press).
11. Owen, T. et al. *J. geophys. Res.* **82**, 4635 (1977).
12. Clayton, R.N. & Mayeda, T.K. *Earth planet. Sci. Lett.* **62**, 1 (1983).
13. Wiens, R.C. et al. *Earth planet. Sci. Lett.* (in the press).
14. Melosh, H.J. *Geology* **13**, 144 (1985).

R. O. Pepin is in the School of Physics and Astronomy, University of Minnesota, Minneapolis 55455, USA.

Agriculture and environment

Cyanobacteria employed as fertilizers and waste disposers

from Rabindra N. Padhy

AMONG free-living, nitrogen-fixing microorganisms, the filamentous heterocystous cyanobacteria (blue-green algae) make the largest global contribution to biological nitrogen fixation¹. In the Orient, they grow naturally, or are cultured and applied to rice fields as green fertilizers, a practice termed 'algazation'. In several countries with different agronomical conditions and soil types, it has been established that cyanobacteria in rice fields decrease the amount of nitrogen-based chemical fertilizers needed to produce a higher rice yield^{2,4}. Even more effective is the use of *Azolla*, a water fern that has as a symbiont the cyanobacterium *Anabaena azollae*, which fixes about three kilograms of atmospheric nitrogen per hectare per day⁵. But the regular use of pesticides during rice cultivation tends to raise the concentration of residues to levels that might inhibit the growth of cyanobacteria. Recently, pesticide-resistant mutants of nitrogen-fixing cyanobacteria have been isolated^{6,7}. Such mutants hold the promise of increasing the effectiveness of green fertilizers used in rice cultivation, thus reducing the dependence on expensive chemical processes for fertilizer production.

Adverse effects of insecticides, fungicides, herbicides and heavy metals on cyanobacteria have been discussed⁷. Some pesticides do not affect and may even enhance the growth of cyanobacteria in laboratory cultures or in the field. But many pesticides at very low concentrations decrease the total nitrogen content of flasks containing heterocystous cyanobacteria growing under nitrogen-fixing conditions. The effect seems to depend on the structure and concentration of the chemical and the nutritional status of the cyanobacterial growth-medium.

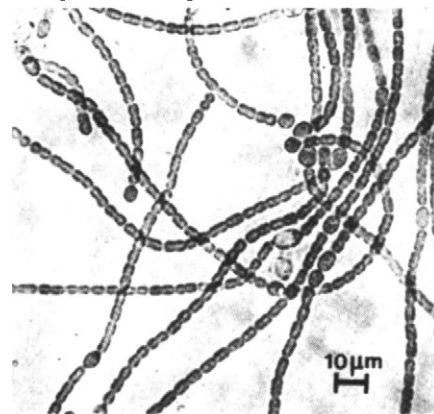
Spontaneous occurrence of pesticide-resistant mutants is low, so a method of inducing mutations is required. A mutagen like *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine is suitable, and mutants of *Nostoc* and *Cylindrospermum* that are resistant to the pesticides blitox, carbaryl, zineb and mancozeb have been isolated^{6,7}. These mutants survive in concentrations of chemicals higher than the highest permissible concentrations for the wild type — that is, the maximum concentration at which cultures contain viable cells. In laboratory cultures, the growth of the mutants in the presence of pesticides is exactly the same as that of the wild strain in absence of the pesticides. Of course, there is as yet no report of the use of such mutants in field experiments.

The heterocystous cyanobacteria that are isolated as spores or filaments from rice fields are not known to have any harmful effect on the environment in general or on animal or human populations in particular, so pesticide-resistant mutants can safely be introduced to the environment. The Food and Agricultural Organization and governments of rice-growing nations in the Orient, Africa, Latin America and United States could consider field trials.

The development of pesticide-resistant strains does, however, have its dangers. Some species of cyanobacteria tend to form surface blooms on inland tropical and temperate waters in the summer, with attendant risks of eutrophication and/or toxin production^{8,9}. It is important that pesticide-resistant strains of these species are not released into the environment, as the blooms they would form might be very difficult to control; in particular the use of mutants of toxin-producing strains must be avoided.

The presence of synthetic and unwanted chemicals released through industrial and urban wastes, and pesticides in soil and water, are matters of environmental concern and studies on biodegradation of the foreign chemicals by microorganisms^{9,10} have been intensely pursued. Not all such chemicals are, however totally biodegradable because microorganisms cannot utilize certain toxic chemicals to proliferate. Nor are degradable chemicals acted on in all types of ecosystems^{9,10}. Such chemicals persist and migrate by leaching. In running waters, they ultimately accumulate near or at the top of food chains. Most studies on biodegradation of synthetic organic chemicals have been carried out with heterotrophic microorganisms and a few purple bacteria.

The studies of C.E. Cerniglia, C. Van Baalen and their colleagues at the University of Texas have indicated that many cyanobacteria originally isolated from both freshwater and marine habitats, metabolize (or even more accurately 'co-metabolize') naphthalene, biphenyl, anilines and their derivatives, and toxic substances present in the water-soluble fraction of crude oils; these are the common foreign chemicals released from chemical plants and dye works^{7,11-13}. The cyanobacterium *Oscillatoria* in particular has several pathways for the catabolism of naphthalene and aniline that are also found in both heterotrophs and eukaryotic algae. Although the detailed mechan-



Photomicrograph of filaments of the cyanobacterium *Anabaena* 7120 (*Nostoc*) showing chains of vegetative cells, the sites of oxygenic photosynthesis, and single larger heterocysts, the sites of nitrogen fixation ($\times 458$).

isms of chemical degradation by different cyanobacteria are not known, probably they are promoted by both the heterotrophic potency and the oxygenic photoautotrophy of cyanobacteria. An advantage of cyanobacteria over eukaryotic algae is the ability to grow in near anaerobic conditions and low light intensities.

All heterotrophic microorganisms face the problem of maintaining a steady level of cellular energy production to support the catabolism required for breaking down these chemicals¹⁰. Furthermore, oxygen is probably the rate-limiting factor in the biodegradation of many aromatic

hydrocarbons by heterotrophs⁹. These problems should not affect cyanobacteria and, if photosynthesis is undisturbed, such co-metabolic processes should continue in alkaline environments.

Normally, cyanobacteria are not the dominant constituent of phytoplankton and in natural ecosystems waste chemicals can be so toxic to their natural populations that particular species are eliminated. But mutant cyanobacteria resistant to the chemicals of interest, once isolated, could be exploited profitably for breaking down the chemicals in the waste-water stabilization ponds that receive industrial and urban effluents. Industries that release recalcitrant molecules in their effluents might consider trials in model ecosystems and the United Nations environment programme could verify the possibility in field extension studies.

Cyanobacteria are found in a wide range of habitats, ranging from ice fields to rice fields, hot springs and sewage. They were the first oxygenic photoautotrophs to convert the reducing atmosphere of the archaic Earth to its present oxidized state. A bloom-forming species, *Anabaena flosaquae*, is reported to grow well in the laboratory in a medium contaminated with oil leached from large oil-shale reserves¹⁴. Are these anabolically versatile organisms, which combine oxygenic photosynthesis with oxygen-sensitive nitrogen fixing systems, also catabolically potent?

Chimaeric plasmids acting as shuttle vectors between unicellular or filament-

ous cyanobacteria and *Escherichia coli* have already been constructed, showing that foreign genes can function in cyanobacterial cells¹⁵⁻¹⁷. With modern recombinant DNA technology, the development of novel strains of cyanobacteria capable of degrading a variety of chemicals should be feasible, by cloning and expression of eubacterial chemical-degrading plasmids from aerobic and anaerobic bacteria that are otherwise environmentally less successful. □

1. Mishustin, E.N. & Shil'nikova, V.K. *Biological Fixation of Atmospheric Nitrogen* (Macmillan, London, 1971).
2. Venkataraman, G.S. *Algal Bio-fertilizers and Rice Cultivation* (Today and Tomorrow, New Delhi, 1972).
3. Fogg, G.E. in *Algal Physiology and Biochemistry* (ed. Stewart, W.D.P.) 574 (Blackwell, Oxford, 1974).
4. Ricer, P.A. & Kulasooriya, S.A. *Blue-green Algae and Rice* (International Rice Research Institute, Manila, 1980).
5. Swaminathan, M.S. *Sci. Am.* **250** (1), 62 (1984).
6. Vaishampayan, A. & Prasad, A.B. *Envir. exp. Bot.* **22**, 427 (1982).
7. Padhy, R.N. *Residue Rev.* **95**, 1 (1985).
8. Reynolds, C.S. & Walsby, A.E. *Biol. Rev.* **50**, 437 (1975).
9. Atlas, R.M. *Microbiol. Rev.* **45**, 180 (1981).
10. Alexander, M. *Science* **211**, 132 (1981).
11. Cerniglia, C.E., Van Baalen, C. & Gibson, D.T. *Arch. Microbiol.* **125**, 203 (1980).
12. Cerniglia, C.E., Van Baalen, C. & Gibson, D.T. *J. gen. Microbiol.* **116**, 495 (1980).
13. Cerniglia, C.E., Freeman, J.P. & Van Baalen, C. *Arch. Microbiol.* **130**, 272 (1981).
14. McKnight, D.M., Pereira, W.E., Rostad, C.E. & Stiles, E.A. *Bull. envir. Contam. Toxic.* **30**, 6 (1983).
15. Van den Hondel, C.A.M.J.J. et al. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1570 (1980).
16. Buzby, J.S., Porter, R.D. & Stevens, S.E., Jr. *J. Bact.* **154**, 1446 (1983).
17. Wolk, C.P. et al. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1561 (1984).

Rabindra N. Padhy of Kahlilikote College, Berhampur, Orissa, India is at present in the Unit of Molecular Genetics, International Institute of Cellular and Molecular Pathology, Avenue Hippocrate 75, B-1200 Brussels, Belgium.

Electron trapping

Hysteretic relativistic resonance of a single electron

from A. E. Kaplan

A SINGLE electron, trapped in a volume of less than one cubic millimetre for an unprecedented 10-month term of confinement, has allowed the first observation of a predicted fundamental hysteretic effect¹. It is due to the relativistic increase — albeit very slight — of mass with velocity. In a recent issue of *Physical Review Letters*, G. Gabrielse and others at the University of Washington, Seattle, report that the cyclotron resonance of a single electron trapped in a Penning trap was hysteretic². The striking fact is that the relativistic change required^{1,2} is very small; since the excitation energy of the electron was usually less than 1 eV, the mass change was of the order of only a millionth of the electron rest mass, m_e .

Why does the hysteresis occur? In a magnetic field B , an electron rotates with a cyclotron frequency $\omega_c = eB/mc$ (where e is the electrical charge of an electron and c is the speed of light), with its mass m relativistically dependent on its velocity v .

Therefore, when the electron is excited by the resonant driving field (with its driving frequency ω close to ω_c), the cyclotron frequency decreases. Suppose now that the initial driving frequency of the external field is larger than the non-relativistic cyclotron frequency ω_0 (corresponding to m_e) and is slowly swept downwards, towards the exact resonance. As the driving frequency approaches the resonant frequency, the electron gets excited and its dimensionless orbit speed $\beta = v/c$ increases. If relativistic effects could be ignored, β would show the familiar Lorentzian resonant response. If, however, the amplitude E of the driving field is sufficiently large, the relativistic redshift of the cyclotron frequency yields a drastically different response, which is depicted by the solid curve in the figure. As the driving frequency ω is swept downward, dimensionless speed β and kinetic energy $\beta^2/2$ increase (upper branch of the solid curve in the figure), and the cyclotron frequency

ω_c is shifted further away from the unperturbed resonant frequency ω_0 as if it is running away from the driving frequency.

This continues until the maximum level of excitation β_m is reached at the point of exact resonance, where $\omega = \omega_c = \omega_0$ ($1 - \beta_m^2/2$). The maximum level of excitation β_m is proportional to the amplitude of a driving field and inversely proportional to the frequency width of resonance Γ , which is determined by the radiation dissipation of energy. If ω is swept further downwards, the speed of the excited electron begins to diminish, and the effective cyclotron frequency starts to run back. For this reason, very shortly after the driving frequency passes the point of exact resonance, the excitation can no longer be supported at the high level of the resonant state and suddenly jumps to the off-resonant state. If the driving frequency is then swept backwards, the other jump (from lower to upper branch in the figure) occurs when ω is very close to ω_0 .

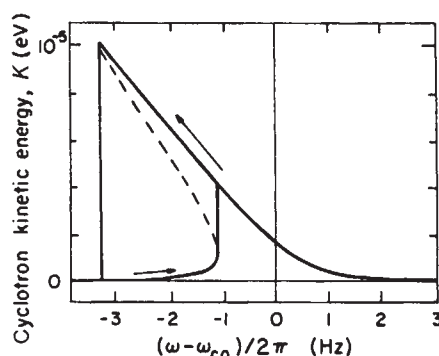
This cycle forms a relativistic hysteresis within the range in which the electron can have two possible steady states of excitation — the simplest and probably the most fundamental bistable system. (The middle branch in the figure corresponds to an unstable regime¹, and cannot be observed.) For the hysteresis to occur, the maximum relativistic shift of the cyclotron frequency, $\omega_0 \beta_m^2/2$ must be sufficiently greater than (but of the same order of magnitude as) the width of the resonance, Γ . This condition yields¹ an unexpectedly low critical energy of excitation for the electron that is nearly α^{-1} times smaller than a quantum limit $\hbar\omega$ ($\alpha = e^2/\hbar c = 1/137$ is the fine-structure constant). Therefore, in the close vicinity of the threshold, only the quantum approach can give an adequate description of the phenomenon whereas, for a sufficiently strong driving field, the classical results (in particular, hysteretic jumps) remain valid. For the experimental situation the critical kinetic energy of excitation is of the order of 10^{-5} eV but the actual excitation used by Gabrielse *et al.* was a few orders of magnitude greater and so the experimental data obviously represent the classical limit.

The fact that the frequency of cyclotron motion decreases as kinetic energy of the particle increases is well known in the theory of the application of synchrotrons and synchro-cyclotrons to the study of ultra-relativistic particles. For Gabrielse *et al.* to observe cyclotron hysteresis it was necessary to work at energies as low as a few electron volts, which only became possible with improved measurements of single particles in the Penning trap³. (A single-atom trapping technique has simultaneously been developed based on laser cooling⁴.) The Penning trap is essentially a positive ring and two negative cap electrodes kept at liquid helium temperature in the core of a 60-kG superconducting magnet. The electron was prevented from

escaping by confinement in a relatively weak quadrupole potential of the trap. The potential of the trap was also used to measure a relativistic mass increase through the axial frequency shift.

Apart from one basic difference, the observed hysteresis very much resembles the so-called vibration hysteresis in anharmonic oscillators with nonparabolic potential¹, in particular the Duffing oscillator in which the shift of resonant frequency is due to the higher-order terms of potential. The difference is that both factors responsible for the single-electron hysteretic effect are of fundamental nature: the relativistic change of mass (and hence cyclotron frequency), and a damping factor Γ which depends only on basic constants.

That is why the observed effect is so important. It may be viewed as the ultimate example of optical bistability (reviewed in ref. 6), which is a rapidly expanding and promising field in nonlinear



Resonant line shape (solid curve) of a slightly relativistic cyclotron resonance. The localization of the electron's steady-state kinetic energy on either upper or lower branch of the solid curve depends upon the direction (arrows) in which the driving frequency is swept through resonance and includes discontinuous jumps (broken line).

optics that offers new insights into the interaction of light with matter as well as the potential for superfast switching devices for optical computers and signal processing. The observed effect is also the first example of an intrinsic optical bistability; all other optical bistable effects depend upon a change in the macroscopic state of a nonlinear medium (the so-called optical feedback).

What of the future? The University of Washington group is planning to measure the magnetic moment of an electron with much improved accuracy using hysteretic resonance. Obviously, it will be of great interest to study the hysteresis in the vicinity of its threshold since this will provide a unique opportunity to study a quantum limit of bistable oscillators. Our group is working on this kind of quantum theory of the phenomenon as well as new multiphoton methods of optical excitation of single-electron cyclotron resonance. As far as the classical limit is concerned, one may envisage that, under the action of a sufficiently strong driving field, the upper

branch of the hysteretic curve will become unstable, resulting in the chaotic motion of an electron similar to chaos in other nonlinear oscillators. Finally, it is worth noting that a similar hysteretic behaviour might be observed for electrons in semiconductors⁷, in which case the relativistic-like change of the effective mass of conduction electrons would be attributable to the non-parabolic nature of the semiconductor conduction band. □

Molecular genetics

What turns on heat shock genes?

from Sean Munro and Hugh Pelham

THE 'heat-shock' response involves a massive increase in the transcription of a small family of heat-shock genes, so it provides a good model system for studying gene regulation in organisms as diverse as *Escherichia coli* and humans¹⁻³. One of its most interesting aspects is the communication between the environment and the genetic apparatus: somehow, the cell detects a rise in the ambient temperature, or one of a number of other stresses, and rapidly and specifically alters its pattern of gene expression. Recent results suggest that the system of ubiquitin-linked protein degradation has a key role in this communication.

Initial work focused on the DNA sequences required for the transcriptional activation of heat-shock genes. Deletion analysis of the *Drosophila hsp70* promoter gene identified a short sequence that is required for induction⁴, which was subsequently shown to be the binding site for a specific heat-shock transcription factor⁵. This factor is present in cells before heat shock, but nuclease-protection studies of *Drosophila* chromatin show that it does not bind to the genes until a stress is applied⁶. The obvious implication is that the availability and/or activity of this factor is modified in some way as a result of the inducing stress. What is less obvious is the nature of the modification and of the intracellular signal that initiates it — how does the cell know that it is stressed?

The heat-shock response can be induced by many treatments other than heat, most, if not all, of which have the common property of causing the accumulation of denatured or damaged proteins within the cell. As it seems unlikely that all the treatments affect a single 'sensor' protein in the same way, it has been suggested that the presence of abnormal proteins *per se* triggers the response^{7,8}. If this is so, then overproduction of one abnormal protein should suffice, which seems to be the case. Certain flightless *Drosophila* mutants make a truncated actin in their flight muscles that disrupts cellular structures and causes expression of heat-shock proteins in those cells^{9,10}. Similarly, heat-shock proteins can be induced in *E. coli* by high-

1. Kaplan, A. E. *Phys. Rev. Lett.* **48**, 138 (1982).
2. Gabrielse, G., Dehmelt, H. & Kells, W. *Phys. Rev. Lett.* **54**, 537 (1985).
3. Gabrielse, G. & Dehmelt, H. in *Precision Measurements and Fundamental Constants II* (ed. Taylor B. N. & Phillips, W. D.) (US NBS Spec. Publ. 617, Washington, DC, 1984).
4. Petley, B. W. *Nature News and Views* **315**, 716 (1985).
5. Landau, L. D. & Lifshitz, E. M. *Mechanics* (Pergamon, New York, 1976).
6. Abraham, E. & Smith, D. *Rep. Prog. Phys.* **45**, 815 (1982).
7. Kaplan, A. E. & Elci, A. *Phys. Rev.* **B29**, 820 (1984).

A. E. Kaplan is Professor of Electrical Engineering, Purdue University, West Lafayette, Indiana 47907, USA.

level expression of human tissue plasminogen activator, which fails to fold correctly in this system⁷.

Ubiquitin

Normally, aberrant proteins are simply degraded by the cell. In eukaryotic cells, they are recognized by the ubiquitin-dependent degradation system; they are covalently attached by the N-terminal (and lysine) amino groups to ubiquitin, a 76-amino-acid protein, through a peptide bond, to become substrates for proteolysis (see ref. 9 for review). After heat shock, however, abnormal proteins fail to be degraded efficiently, which can easily be demonstrated for individual truncated proteins expressed from plasmids transfected into mammalian cells¹¹. The degradation system itself, assayed *in vitro*, is not particularly temperature sensitive; it seems, rather, to be overloaded by the presence of high levels of heat-denatured protein. We observe the same failure to degrade abnormal proteins when cells are treated with other inducers of the stress response, such as sodium arsenite and amino-acid analogues, with a close correlation between the two phenomena. This strongly suggests that the treatments commonly used to induce the heat-shock response do indeed cause abnormal protein to accumulate.

Perhaps the strongest evidence for a connection between the degradation system and the stress response comes from studies of the ts85 mouse cell line. Cells of this line have a thermolabile ubiquitin-activating enzyme; at the non-permissive temperature, which is well below the normal heat-shock threshold, ubiquitin can no longer be coupled to proteins and degradation ceases. At the same time, the cells synthesize heat-shock proteins at a high level⁸. This may reflect a build up of abnormal proteins, but it could also be a more direct effect of the failure to attach ubiquitin to proteins.

In fact, several lines of evidence suggest that it is ubiquitination, rather than proteolysis, that becomes limiting in normal eukaryotic cells during stress. First, there are dramatic changes in the levels of

100 years ago

The St. Andrews Marine Laboratory

PROF. M'Intosh stated briefly the structure and arrangement of the marine laboratory at St. Andrews, and made some general remarks on the work done during the last nine months there. A great many of our food fishes, he said, were carefully examined in regard to the development of the eggs and the growth of the young fishes. About twenty species were examined in this way. They experienced some difficulty with some of the forms, on account of their voracity, particularly with the cod. They found that a cod of five inches long would swal-

low a cod of three inches, and if it could not get it all down at once, it would keep it in its throat till the head part was digested, and then draw in the tail. Mollusca were studied chiefly in connection with the development of the mussel, but he might say that very hazy notions were held in regard to it. Some larger forms were also examined, including porpoises and sharks. He had noticed one porpoise for some time in the bay, and that its motions were peculiar. Then a large porpoise was caught, and they found it was a female giving milk, of a most interesting kind, as dense as cream, and a deep yellow colour.

from *Nature* 32 563, 8 October 1885.

ubiquitinated histones H2A and H2B. About 10 per cent of H2A normally has a ubiquitin moiety coupled to the ϵ -amino group of a lysine residue; the function of this modification is unclear, but the protein is not rapidly degraded. The attached ubiquitin is in rapid equilibrium with the pool of free ubiquitin, being continually removed by an isopeptidase (see ref. 8 for review). When *Drosophila* cells are heat-shocked the level of ubiquitinated histone is greatly reduced¹², presumably because the isopeptidase remains active but ubiquitin is no longer replaced. Similarly, the abnormal proteins that accumulate to high levels during the heat shock of mammalian cells are almost entirely in the non-ubiquitinated state, as judged by the size of specific proteins on SDS gels (ref. 11 and unpublished observations), which suggests that the point of inhibition of the degradative pathway is at, or before, the addition of ubiquitin.

The ubiquitin gene itself has recently been cloned from several species^{13,14}, and strikingly, its transcription was found to be strongly heat-inducible in both chick fibroblasts¹⁴ and yeast. Presumably this reflects an increased need for it after heat shock, which in turn suggests that the free ubiquitin level may become limiting. These observations are all consistent with the idea that a sudden heat shock produces such a massive increase in substrates requiring ubiquitination that there is a transient shortage of free ubiquitin, resulting in a failure to ubiquitinate normal substrates such as histones.

A model

Could it be that lack of free ubiquitin is the signal initiating the heat-shock response? Or is it the failure to couple ubiquitin to protein? The properties of ts85 cells argue against the first possibility. At the non-permissive temperature, they are unable to couple ubiquitin to anything, and thus presumably have high levels of free ubiquitin, yet they synthesize heat-shock proteins. However, there is support for the alternative, that the heat-shock response occurs because some target protein fails to be ubiquitinated. An obvious candidate is the heat-shock transcription factor itself. We speculate that, by analogy with the H2 histones, the factor is normally maintained in an inactive ubiquitinated form, but with the ubiquitin moiety

continually being removed and replaced. Treatments of the cell that either inhibit the ubiquitinating enzymes or produce a large sink of abnormal protein substrate would allow accumulation of non-ubiquitinated active factor, which in turn would promote transcription of the heat-shock genes. Such a mechanism would be inherently self-regulating: as soon as free ubiquitin becomes available, either through degradation of the abnormal protein or through increased synthesis of ubiquitin, the transcription factor would again be inactivated. This could explain the old observation that transcription of the heat-shock genes occurs only transiently, even when cells are maintained at a high temperature for long periods¹.

Recent studies of the factor are consistent with this model. At a meeting* earlier this year, Carl Parker reported the purification of the protein from heat-shocked *Drosophila* cells; the active fraction included a protein with an apparent relative molecular mass on SDS gels of 43,000 (43K). Tantalizingly, the equivalent fraction from unshocked cells included small

amounts of this 43K protein and a related 48K protein, as well as other minor proteins. It is conceivable that these are the transcription factor and its ubiquitinated derivative, but this is far from proven.

To test the model, it will be necessary to demonstrate that the transcription factor is ubiquitinated and that the ubiquitinated form is not bound to heat-shock promoters *in vivo*. A more general prediction is that the heat-shock proteins induced in response to protein damage have functions which help to limit or repair such damage, for example, by binding to exposed hydrophobic surfaces¹⁵. The major heat-shock protein, hsp70, certainly seems to aid in the repair of heat-damaged nucleoli¹⁵.

Many of the proteins involved in cell transformation and growth control, such as *c-fos*, *c-myc* and adenovirus E1A proteins are 'abnormal' in being very unstable and presumably substrates for ubiquitin ligases. It is not inconceivable that ubiquitination plays a regulatory role in areas other than the heat-shock response. □

1. Ashburner, M. & Bonner, J.J. *Cell* **17**, 241 (1979).
2. Schlesinger, M.J., Ashburner, M. & Tissieres, A. eds *Heat Shock from Bacteria to Man* (Cold Spring Harbor, 1982).
3. Pelham, H.R.B. *Trends Genet.* **1**, 31 (1985).
4. Pelham H.R.B. & Bienz, M. *EMBO J.* **1**, 1473 (1982).
5. Parker, C.S. & Topol, J. *Cell* **37**, 273 (1984).
6. Wu, C. *Nature* **309**, 229 (1984).
7. Goff, S.A. & Goldberg, A.L. *Cell* **41**, 587 (1985).
8. Finley, D. et al. *Cell* **37**, 43 (1984).
9. Karlik, C.C., Coutu, M.D. & Fyrberg, E.A. *Cell* **38**, 711 (1984).
10. Hiromi, Y. & Hotta, Y. *EMBO J.* **4**, 1681 (1985).
11. Munro, S. & Pelham, H.R.B. *EMBO J.* **3**, 3087 (1984).
12. Glover, G.V.C. in *Heat Shock from Bacteria to Man* (eds Schlesinger, M.J., Ashburner, M. & Tissieres, A.) (Cold Spring Harbor, 1982).
13. Ozkaynak, E., Finley, D. & Varshavsky, A. *Nature* **312**, 663 (1984).
14. Bond, U. & Schlesinger, M.J. *Mol. Cell Biol.* **5**, 949 (1985).
15. Pelham, H.R.B. *EMBO J.* **3**, 3095 (1984).

Sean Munro and Hugh Pelham are at the Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, UK.

Geophysics

Small-scale mantle convection

from Luce Fleitout

DURING the last few years, the tectonic evolution of basins has attracted much attention mainly because of the potential importance for oil exploration. Hydrocarbon formation depends critically on the rate of sedimentation and the temperature within the sedimentary layers, both of which are greatly affected by the thermal history of the basin. There are two main tectonic processes which can modify the surface topography of the Earth; a global thinning (or thickening) of the entire lithosphere, and thermal perturbations of the lower lithosphere without deformation of the upper crust. In a recent issue of *Nature*, M. S. Steckler¹ demonstrates that the Gulf of Suez was formed by global lithospheric thinning followed by additional warming up of the lower lithosphere. The mechanism probably respon-

sible for this increase of temperature is a small-scale convective destabilization involving the base of the lithosphere. This has implications which extend beyond merely understanding better the geology of the Gulf of Suez, since it confirms the importance of small-scale convection in tectonic processes.

The geology and geophysics of a number of sedimentary basins has been successfully explained by a model of uniform stretching of the lithosphere. The lithosphere (Fig. 1) consists of the crust and a portion of the upper mantle. The crust is less dense than the underlying mantle, whereas the lower lithosphere is cooler and hence denser and more viscous than the sublithospheric mantle. When no process other than stretching occurs, both the crust and the cold lithosphere are thinned

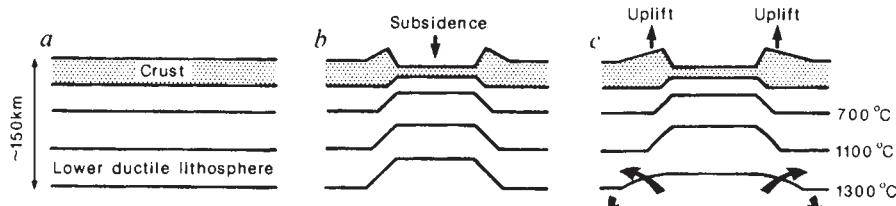


Fig. 1. Model of uniform stretching of the lithosphere to form sedimentary basins. *a.* Undeformed lithosphere comprising crust and cold mantle material. *b.* Stretched lithosphere — both the crust and the uppermost mantle are thinned and isostatic adjustment induces a topographic depression. *c.* Destabilization by a convective process, of the lower lithosphere induces uplift. Topographic effects have been exaggerated. Solid lines represent mantle isotherms.

in that the total extension is the same for both components. The thinning of the crust induces a depression by local isostatic adjustment. The thinning of the lower lithosphere is accompanied by an uplift extending over an area that can be broader than this depression, because compensation of deep mass anomalies is not necessarily local and also because stretching of the lower lithosphere can be less concentrated than stretching of the crust. The integrated area encompassing the topographical anomaly curve (that is, the horizontally integrated depression of the rift minus the integrated uplift of the shoulders) is uniquely determined once the total extension and the initial crustal thickness are estimated, as it obeys regional isostasy. Following the phase of tectonic deformation, sediments fill the central depression and long-term subsidence takes place, due to the conductive cooling of the stretched lithosphere.

In the particular case of the Gulf of Suez, Steckler shows that the observed topography does not correspond to this classical picture, the sides of the rift being higher than expected. The very large uplift of the rift flanks implies an extra warming-up of the lower lithosphere. This idea is not entirely new: the differential stretching found necessary in the Pannonian basin³ could, for example, result from the action of the subduction event in the nearby Carpathians. But the Gulf of Suez is far away from other zones of active tectonism and Steckler is able to rule out most of the other mechanisms that could be invoked to explain the extra heating of the lithosphere required. Destabilization of the lower portion of the lithosphere by small-scale convection (Fig. 1c) is the most likely hypothesis.

What is the cause of this destabilization of the lower lithosphere? The viscosity of the mantle is highly temperature dependent, so the cold upper part of the lithosphere is unable to flow under moderate stresses and only the lowermost and warmest portions can deform plastically. Even so the material forming the lower part of the lithosphere is colder and denser than the underlying mantle and thus tends to sink under its own weight. The lower the viscosity of the lower mantle, the faster the convection.

In the case studied by Steckler, a hori-

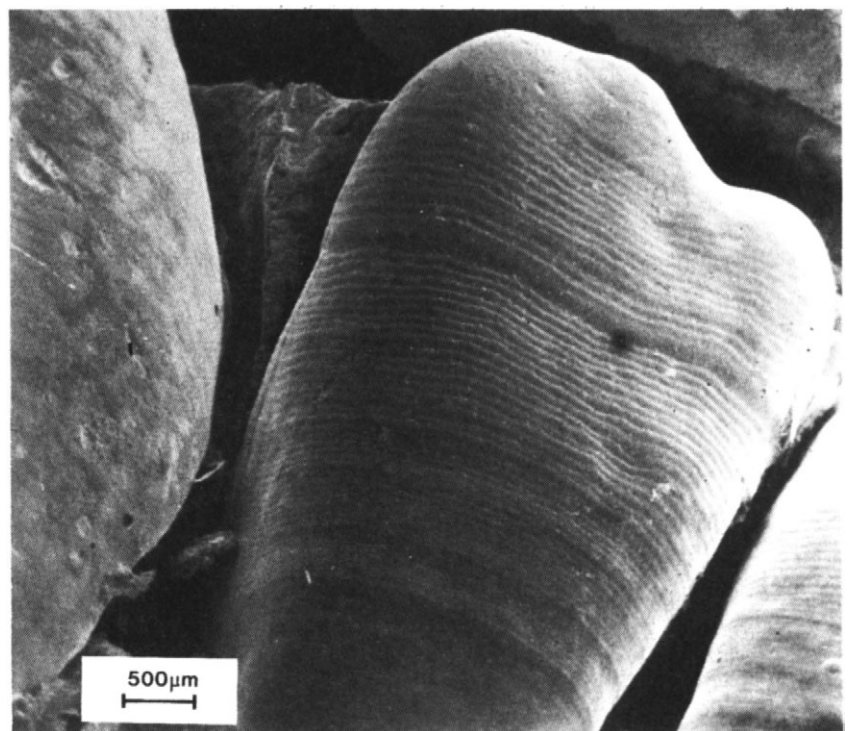
zontal gradient of temperature enhances the phenomenon of destabilization and is able to produce a large degree of uplift on the flanks of rifts. Other regions close to a passive continental margin, such as Scandinavia or eastern Australia, may also have been uplifted in a similar fashion after the opening of the adjacent sea.

But convective processes at the lithosphere-asthenosphere boundary can be important in a number of other situations. During continent-continent collision, as for example in the Alps or the Himalayas, the whole lithosphere thickens in an exact reversal of its behaviour during basin formation. The thickened cold part of the lithosphere plays a significant role in further enhancing the compression beneath the mountain ranges^{4,5}, as has been

revealed by seismic tomography, both in the Alps⁶ and in California⁷. Destabilization and sinking of this cold blob in the mantle can terminate the compressive phase and be accompanied by topographic uplift. In essence, the whole evolution of the lithosphere seems to be dominated by the convective interaction between the lithosphere and the underlying mantle. Small-scale convection can be thought of as a continuing convective flow pulling the cold material at the base of the lithosphere. It supplies heat to the lithosphere-asthenosphere boundary. This process is able to govern the rate of thickening of the lithosphere in the oceans, to fix the asymptotic lithospheric thickness under old cratons^{8,9} or to produce rapid thinning of the lithosphere over zones of high temperature and low viscosity associated with hot spots. □

1. Steckler, M.S. *Nature*, **317**, 135 (1985).
2. McKenzie, D. *Earth planet. Sci. Lett.* **40**, 25 (1978).
3. Sclater, J.G. *et al. Earth planet. Sci. Lett.* **51**, 139 (1980).
4. McKenzie, D. *Geophys. J. R. astr. Soc.* **55**, 217 (1978).
5. Fleitout, L. & Froidevaux, C. *Tectonics* **1**, 21 (1982).
6. Babuska, V., Plomerova, J. & Sileny, J. *Ann. Geophysicae* **2**, 649 (1984).
7. Humphreys, E., Clayton, R.W. & Hager, B. *Geophys. Res. Lett.* **7**, 625 (1984).
8. Fleitout, L. & Yuen, D.A. *Phys. Earth Planet. Int.* **36**, 181 (1984).
9. Buck, R. & Parmentier, M. *J. geophys. Res.* (in the press).

Luce Fleitout is in the Laboratoire de Géophysique et Géodynamique Interne, Université de Paris-Sud, 91405 Orsay, France.



New ages for fossil hominids from their teeth

Estimating the ages at which fossil hominids died should be more accurate in future, using observations reported by T. G. Bromage and M. C. Dean on page 525 of this issue. This scanning electron micrograph shows the buccal surface of a modern human permanent lower incisor with the perikymata (Greek: *peri* = around; *kymata* = waves) that are the external ends of striations resulting from a periodicity of seven to eight days in the deposition of enamel. The perikymata are used to obtain accurate estimates of the ages at death of four species of fossil hominids.

Controversial glycosaminoglycan conformations

SIR—Despite major recent advances in the conformational analysis of glycosaminoglycans, several controversies remain concerning the interpretation of physical and chemical measurements. One major issue concerns the ring form of L iduronate residues in a number of important glycans and proteoglycans having dermatan, heparin or heparan chains. At least for dermatan sulphate, the evidence from diffraction analysis¹ is that these residues in crystalline fibres are close to the form having the carboxylate group axial (⁴C₁, Fig. 1). In solution, however, this would not preclude the alternative chair (¹C₄, Fig. 1) since it is well known that the interaction energies are such that equilibrium between the two forms is delicately balanced² and there is of course no reason *a priori* why the crystalline form should persist into solution. Also possible are various distorted chairs and skew boats but these are not discussed here since their consideration would constitute a second order of analysis.

Two conflicting arguments have been put forward. On the one hand, the small proton nuclear magnetic resonance vicinal coupling constraints around the ring would point³ to ¹C₄, an assignment that is confirmed by our recent studies by circular dichroism⁴. On the other hand^{5,6}, the extreme susceptibility of these residues to periodate oxidation must be taken as a demonstration of an ability to form a cyclic reaction intermediate which requires a small dihedral angle between the 0(2) and 0(3) substituents, as in the ⁴C₁ (Fig. 1). We now write to point out that this apparent discrepancy can be resolved by the application of established stereochemical principles which show that the periodate

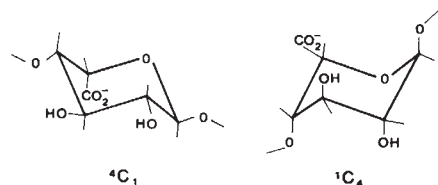


Fig. 1 Alternative chair conformations for L-iduronate, with designations as used in text.

oxidation behaviour can be understood without the need to postulate a preponderance of ⁴C₁ in the equilibrium between ground state conformations.

It follows from the Curtin-Hammett principle that — when alternative conformers of a reactive compound are in rapid equilibrium relative to the rates of their subsequent reaction — the reaction pathway is independent of their relative ground state energy levels (abundancies); thus a major conformer (in this case ¹C₄) can function as an unreactive reservoir species while a very minor conformer (⁴C₁) determines the course of reaction. Another very striking example of this principle has been described recently dur-

ing the catalytic asymmetric hydrogenation of a prochiral dehydro-amino acid where the reactive intermediate of a rapidly equilibrating pair was actually undetectable. This leads to the postulate that the iduronate and glucuronate residues are oxidized in similar conformations, in which case one might at first expect the latter to react more quickly because of the greater probability of productive collisions — whereas the reverse might be the case. This problem can be resolved by reference to a kinetic study of the cyclohexane-1, 2-diols which provides evidence that the release of crowding in their cyclic intermediates in periodate oxidation can provide the driving force for steric acceleration, at least at pH 9 (ref. 9). We suggest that this indicates that transition states of this type are sufficiently product-like in character (that is, C-C bond cleavage is well progressed) for steric crowding to be substantially relieved. The energy levels for the transition states for the glucuronate and iduronate pathways would then be closer in energy than those for their ground state conformations. The higher ground state energy of iduronate relative to that of glucuronate — as a result of the many unfavourable internal interactions — would therefore provide steric acceleration and hence explain the fast rate that is observed for iduronate.

D.A. REES

National Institute for Medical Research,
The Ridgeway, Mill Hill,
London NW7 1AA, UK

E.R. MORRIS

Department of Food Research
and Technology,
Cranfield Institute of Technology,
Silsoe College, Silsoe,
Bedford MK45 4DT, UK

J.F. STODDART

Department of Chemistry,
University of Sheffield,
Sheffield S3 7HF, UK

E.S. STEVENS

Department of Chemistry
State University of New York,
Binghamton, New York 13901, USA

1. Mitra, A.K., Arnott, S., Isaac, D.H. & Atkins, E.D.T. *J. molec. Biol.* **169**, 873–901 (1983).
2. Eliel, E.L., Allinger, N.L., Angyal, S.J. & Morrison, G.A. *Conformational Analysis*, 370 (Wiley, New York, 1965).
3. Perlín, A.S., Casu, B., Sanderson, G.R. & Tse, J. *Carbohydrate Res.* **21**, 123–132 (1972).
4. Csiner, D.G., Stevens, E.S., Morris, E.R. & Rees, D.A. *J. Am. chem. Soc.* (in the press).
5. Scott, J.E. & Tigwell, M.J. *Biochem. J.* **173**, 103–114 (1978).
6. Fransson, L.-Å. *Carbohydrate Res.* **62**, 235–244 (1978).
7. Eliel, E.L. *Stereochemistry of Carbon Compounds* 151–156 (McGraw Hill, New York, 1962).
8. Halpern, J. *Pure appl. Chem.* **55**, 99–106 (1983).
9. Buist, G.J., Bunton, C.A. & Miles, J.H. *J. C. Soc.*, 743–748 (1959).

Human lymphocytopathic retroviruses (HLRV)?

SIR—Owing to the independent discovery of human lymphocytopathic viruses (HLV), called human T-cell leukaemia viruses (HTLV-I, II, III) and, later, human T-cell lymphotropic viruses by Gallo and colleagues at the US National Institute of Health^{1,2}, lymphadenopathy-associated

viruses (LAV) by Montagnier and colleagues at the Institut Pasteur in Paris; or AIDS-associated retroviruses (ARV) by Levy and associates in San Francisco³ — there has been lively discussion^{1–3} with respect to naming the heterogenous retroviridae causing analogous spectra of devastating diseases in humans, cats, monkeys, cattle and mice. Conforming with nomenclature in the other mammals and considering heterogeneity, as well as multiplicity, HLV (in the plural) might constitute a comprehensive abbreviation to connote the leukaemogenic, lymphotropic, lymphocytotropic, lymphocytostatic, lymphocytotoxic, sometimes lymphocytolytic nature of such insidious retroviruses accustomed to use human lymphocytes for reproduction, as well as transport through lymph, blood and other secretions^{4,5}.

However, HLRV might prove to be a better abbreviation for human lymphocytopathic retroviruses because:

(1) H, standing for human, can be replaced with Fe, S, B, Mu or any other suitable symbols to designate other creatures ranging from aardvarks to zebras infected with similar retroviridae causing analogous diseases.

(2) T- might be misleading, since lymphocytopathic retroviridae may infect “B-cells”, as well as other cells in sundry species. So, after Lord Admiral Horatio Nelson at Trafalgar, one might X the “T”.

(3) L related to lymphocytopathies, whether the retroviral effects are leukaemic, lymphomatous, lymphoproliferative, lymphocytostatic, lymphocytolytic, lymphocytopenic or otherwise lymphocytotropic.

(4) A for AIDS is not comprehensive, since the lymphocytopathic retroviridae may alternatively or simultaneously cause GLS or lymphomas. As an abbreviation for associated, A is superfluous. So, one might X the “A” similarly.

(5) R for retro denotes DNA retrotranscription by retroviridae which lack double helices and, thus, depend for replication on cells with relatively great DNA concentrations, especially lymphocytes.

(6) V for viruses must stand, especially if preceded by retro-.

(7) While HTLAV sequentially connotes the works of Drs. Gallo, Montagnier and Levy, the abbreviation is a bit awkward. After the K (Köchel) subclassified works of Wolfgang Amadeus Mozart, it might prove efficient to suffix the HLRV with Arabic numbers 1, 2, 3, *et cetera*, *ad infinitum* (especially because it appears now that too many different antigenic strains will be found to use roman numerals conveniently).

JACK W. SHIELDS
Santa Barbara Medical Foundation Clinic,
PO Box 1200, Santa Barbara,
California 93102, USA

1. Wong-Staal, F. & Gallo, R.C. *Blood* **65**, 253–263 (1985).
2. Wong-Staal, F. *Nature* **314**, 574 (1985).
3. Marx, J.L. *Science* **227**, 1449–1451 (1985).
4. Shields, J.W. *Blood* **62** (Suppl 1), 117a (1983).
5. Olsen G.P. & Shields J.W. *Nature* **309**, 116–117 (1984).

The improvement movement

Michael Lockwood

In the Name of Eugenics: Genetics and the Uses of Human Heredity. By Daniel J. Kevles. Knopf: 1985. Pp. 426. \$22.95.

LET us be clear: this is not, and does not claim to be, a comprehensive history of eugenics. Were such a history to be written today, it would doubtless begin with Plato who, in his *Republic*, advocated rigid state control of all matings as between members of the Guardian caste; and it would end, perhaps, with the current, ill-starred eugenics programme of the Government of Singapore, launched in 1983 (of which more presently). Neither is referred to by Kevles, whose concerns are both narrower and broader than this. Narrower, inasmuch as his main focus is on developments in Britain and the United States, and these only from the time of Francis Galton, to whom we owe the term "eugenics". Broader, because his history encompasses research into human heredity quite generally, even where there is little or no connection with any eugenic aims.

So Kevles's story is essentially a British and American one. Moreover it is a story in which the United States generally compares unfavourably with Britain, for two reasons. First, in the period with which Kevles deals, British academic research into human heredity has generally far outshone that carried out in the United States. Secondly, the British liberal tradition proved sufficiently robust to withstand the political pressures from eugenicists that, in the United States, resulted in widespread and often compulsory sterilization of criminals, the mentally ill and the so-called "feeble-minded".

In the early years, from 1865 until around 1930, eugenics and the study of human heredity appear to have gone hand in hand. Those most closely involved in the study of human heredity — Francis Galton and Karl Pearson in England, Charles Davenport in the United States — were inspired with a social mission to improve the human stock, or at least to prevent its degeneration. In an era newly exposed to statistics, it was easy to arouse public alarm at the fact that those sectors of the community in which crime, "feeble-mindedness", disease and lack of physical fitness were most in evidence were at the same time those that had the largest families. In the United States, such concerns were focused especially on the black community and the tide of immigrants. The early eugenic movement often provided a cloak of scientific respectability for attitudes rooted in a combination of class and race prejudice and even sexual puritanism. (Kevles points to the widespread misconception that sterilization reduced

libido. He points out also that eugenics provided support to opponents of birth control: for wouldn't the ready availability of contraceptives mainly serve to reduce the fertility of those very same educated classes that, on eugenic grounds, should be encouraged to procreate?)

Why, then, did the movement eventually run out of steam? Clearly, the atrocities

Eugenics 'loveboat' heads for port

SINGAPORE's prime minister, Lee Kuan Yew, is being forced to back down on his controversial programmes to get more graduate women married off and producing babies. "Loveboat" cruises to the Maldives and Japan, luxurious days at the local Club Méditerranée and a series of tea-time dances at the Kasbah Club in Singapore's Mandarin Hotel are among the events which have created a social whirl for women now referred to in the Singaporean press as "spinster graduates". The government paid for improvements in child-care and offered graduate mothers priority access to the best primary schools for their children if they produced three or more babies. The idea of setting up a government-sponsored matchmaking programme for the educated elite followed Lee's

Breeding failure — a news story from New Scientist, 16 May 1985.

carried out by the Nazis in the name of eugenics did much to turn public opinion against it. But Kevles provides convincing evidence that the movement was already in decline by the early 1930s — although sterilization laws continued to be passed throughout the decade. His own analysis is that it was largely discredited by the march of science. More careful research served to demolish the theory that what was variously termed "feeble-mindedness" or "mental deficiency" was a unitary condition with a simple hereditary basis. And the likely efficacy of proposed eugenic measures was severely undermined by the realization (a) that environment played a substantial role in the shaping of the human phenotype; (b) that "regression to the mean" meant that the children of genetically highly endowed parents were likely to be less well endowed than their parents; (c) that sterilizing individuals manifesting undesirable genes would still leave those same genes present as "recessives" in outwardly unaffected (and hence unsterilized) individuals. Not that (b) or (c) would have been widely appreciated by the public at large; but the scientists understood these things. And as the leading authorities on human genetics came increasingly to dismiss traditional euge-

tics as "pseudoscience", so it inevitably lost ground politically.

No doubt this is all true, as far as it goes. But just as important, arguably, is the fact that the eugenics movement has never really had, on the positive side, a coherent and realistic political programme. It is all very well saying that those of superior stock should have more children, while those of inferior stock should have less; but how exactly is one to bring that about in a democracy? Though several of the pioneers of eugenics flirted with the idea of state regulation of marriage, they were generally forced to fall back on the pious and fatuous hope that, with suitable encouragement, people would voluntarily modify their reproductive behaviour along eugenic lines. What Kevles calls "mainline eugenics" has always had an air of dottiness about it. Witness Galton's looking forward to the day when a woman would no more think of accepting a man "without knowing his biologicogenealogical history" than a stockbreeder would take "a sire for his colts or calves . . . without pedigree". Or Shockley's (1971) proposal that the state should provide financial incentives to those of below average intelligence to be voluntarily sterilized, the amount offered being proportional to the number of IQ points by which they fell below 100. Only slightly less farcical are some of the recent activities of the Government of Singapore, whose eugenics programme has included state-organized singles cruises, designed to introduce women graduates to eugenically eligible young men. (An expenditure of £100,000 apparently resulted in only two marriages.)

Kevles expresses the hope that an historical understanding of the eugenic movement may help us "thread a path . . . into the uncharted territory of genetic engineering". Perhaps. (Though that may be a little like looking to the history of the temperance movement for guidance on the contemporary problems of drug addiction.) Curiously, it seems that we have never really succeeded in coming to terms with our own genes — something for which the eugenic movement, with the reaction that it understandably provoked, perhaps deserves a share of the blame. The very suggestion that differences of personality and intellect between individuals might depend substantially on genetic differences nowadays causes outrage in many quarters, especially when it is suggested that statistical differences between ethnic groups or the sexes may be so influenced. I doubt if we have much chance of responding wisely and rationally to the new possibilities that genetic engineering and the new reproductive technologies are opening up, so long as ideological preconceptions prevent people from looking dispassionately at the nature-nurture issue.

Whether or not there is anything we can

usefully learn from the development of Anglo-American eugenics and human genetics, it makes a marvellous story. Kevles has done a prodigious amount of research and has uncovered a wealth of fascinating material. In the first part of the book, however, the author's lack of philosophical sympathy with the central characters, Galton, Pearson and Davenport, is rather too much in evidence. I found myself irritated by his periodic forays into psychoanalysis, as when he speculates that "Galton may well have diverted frustration over his own lack of children into an obsession with the eugenic propagation of Galton-like offspring". And his portrait of Pearson seems somewhat one-sided, if the accounts of other authors are anything to go by. Contrast Kevles's assertions that Pearson was "blinded by eugenic prejudice" and that he "often display[ed] a relentless closed-mindedness" with Helen Walker's assessment (in the *International Encyclopedia of the Social Sciences*): "his first thought was to get to the truth and if intellectually convinced of an error, Pearson was ready to admit it". Questions of personality aside, Kevles hardly conveys adequately the magnitude and scope of Pearson's achievements as a scientist, which were by no means restricted to the study of human heredity and statistical theory.

I have, though, nothing but praise for Kevles's treatment of the "golden age" of British genetics: the age of Ronald Fisher, Lionel Penrose, Julian Huxley, J.B.S. Haldane and Lancelot Hogben. It was a time when the Galton Laboratory in London, in spite of being chronically underfunded, boasted an astonishing concentration of talent and completely transformed the study of human genetics. Perhaps, indeed, it was partly *because* of the lack of funds that they achieved so much: Haldane remarked that, in the absence of equipment, they were forced to *think*. It is an exciting, even inspiring tale. And Kevles tells it beautifully.

The work done at the Galton laid the foundations for all modern research on the subject: research which was eventually to give rise to a new branch of medicine. With it came, first, genetic counselling, then programmes of genetic screening — for sickle-cell in blacks and Tay-Sachs in the Jewish population. Also amniocentesis and, latterly, chorionic biopsy, making it increasingly possible to detect genetic and chromosomal abnormalities in the womb and selectively to abort. The advent of *in vitro* fertilization raises possibilities previously undreamt of. We are still a long way from being able, as Huxley put it, to "take charge of our own evolution". But we are a lot closer than when he said it. And no closer, I fear, to knowing what we shall do with this power when we have it.

Michael Lockwood is Staff Tutor in Philosophy in the Department for External Studies, University of Oxford, Rewley House, 1 Wellington Square, Oxford OX1 2JA, UK.

Insight into Earth's changing face

Joe Cann

New Views on an Old Planet: Continental Drift and the History of the Earth. By Tjeerd H. van Andel. *Cambridge University Press: 1985. Pp.324. £15, \$19.95.*

A CENTURY ago, gentlemen of leisure considered it quite the proper thing to take an active interest in geology. After all, geology had brought about the Huttonian Revolution, which extended the age of the Earth at least a thousand-fold, and evidence from fossils was central to the ideas of Darwin. Geology was seen as a key part of scientific culture.

How different is the situation now! The Huttonian Revolution, which achieved as great a cultural change as the Copernican, is now taken for granted, even in the country of its birth. Where is the monument at Jedburgh, where the signpost to Siccar Point? Moreover geology has since brought about another revolution, so that continents are mobile, sliding over the Earth, splitting and colliding, and separated by dynamically evolving deep-ocean basins, rather than fixed and anchored as the old orthodoxy would have it. So why is geology not again a central science? Is it because of the lack of informed popularization by creative geologists? Certainly much popular geological writing is abysmally inept.

Against this background it is particularly interesting to read Tjeerd van Andel's new book. van Andel is not only a scientist who has contributed important elements to the new geological framework, but also a man of scholarship and wide erudition. Here he sets out to tell one part of the new story: how our ideas of geological history have been changed and how we have gained radical new insights into the changing face of the Earth.

This is only one half of the story, an important point to stress, since the book scarcely touches on geological process, the mechanisms by which geological changes occur. This is perhaps partly because the subject is treated entirely without mathematics, but also, I suspect, because van Andel's main interest lies in Earth history. The book arose from a series of lectures to non-geologists at Stanford, and van Andel does try very hard to use the smallest possible selection of the multifarious and notorious specialist vocabulary of geology.

New edition

• *World Nuclear Directory: A Guide to Organizations and Research Activities in Atomic Energy*, 7th Edn, consultant editor C.W.J. Wilson. The *Directory* includes details of some 2,000 laboratories and establishments worldwide, and is published by Longman/Gale Research. Price is £95, \$180.

There are five main sections, starting with the most recent topic, the last ice age, moving on to continental drift, and building on that to examine climates of the past, before proceeding to look at the earliest stages of Earth history and, finally, the history of life. *New Views on an Old Planet* is well enough written and presented, but its main strength is the depth of insight it shows. A scientist of van Andel's calibre does not bankrupt his knowledge in the writing of a single book, but selects creatively from his considerably greater store. Whatever the educational bargain-hunters of the present day say, a close involvement with research does enrich and enliven teaching (and popularization); that shows clearly here. Perhaps the book will encourage other prominent geologists to attempt similar works, and perhaps this in turn will restore to geology some of its former public understanding.

Joe Cann is J.B. Simpson Professor in the Department of Geology, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK.

Catholic interests

H.W. Paul

Uneasy Genius: The Life and Work of Pierre Duhem. By Stanley L. Jaki. *Martinus Nijhoff: 1985. Pp.472. Dfl.175, \$65.50, £44.50.*

"SHH! Papa is looking for a theorem." His daughter's silence did not help Pierre Duhem (1861–1916) to discover any great theorems, but he did enough good physics to find a couple of equations that share his name with Margules and Gibbs. Less because of his physics than because of his work in the history and philosophy of science, Duhem is one of the best known of the generation of nineteenth-century giants whose immense productivity drives the jet-setters of today to despair. *Nature's* obituary in 1916 hit the right note:

If Duhem did not concentrate his main efforts on the discovery of new phenomena or the measurement and remeasurement of physical constants, he at least played an equally important part in the advancement of our knowledge by evolving order out of chaos, and uniting isolated portions of mathematical physics in the form of a connected and logical theory.

Be this as it may, Duhem's classic work in thermodynamics has been resurrected lately by mandarins of the calibre of Prigogine and Truesdell. A reprinting in 1961 of his *Recherches sur l'Hydrodynamique* (first published in 1903) shows the value of older work on the mechanics of fluids for the engineer when he has to worry about the shock waves produced at supersonic speeds. Contrary to the opinions of Perrin and Langevin, Duhem's anti-atomism and opposition to relativity did not doom him

to interment in the graveyard of dead paradigms. Like Berthelot, his lifelong scientific opponent, Duhem also contributed to the development of the modern theory of explosives. Perhaps because he was kept in Bordeaux for most of his academic life, Duhem did not have a stable of students who worked on his ideas, and this contributed to his scientific oblivion in France.

In his brilliant study *The Origins of Modern Science*, written nearly 30 years ago, Herbert Butterfield paid proper homage to the work of Duhem in laying bare the structure of mediaeval physical thought. From the viewpoint of the anticlerical republican establishment, Duhem's discovery of the genius of mediaeval Catholicism was as unwelcome as his embarrassingly cogent attacks on Berthelot's rotten thermochemistry. At least this was of more ideological significance than Duhemian fulminations on the shaky status of Maxwell's worm-eaten equations. As a right-wing, publicly conspicuous Catholic in the state university, Duhem had no chance of getting a chair in Paris. This probably made little difference to the fortunes of science. At the end of the nineteenth century, a new galaxy of scientists appeared in Paris and the provinces — the Curies, Perrin, Sabatier, Brillouin, Langevin — who were also acceptable ideologically to the Third Republic, an intolerant mistress intent on re-establishing France as a great scientific power. By a perverse logic of events, Duhem was also one of the main contributors to another of the Republic's successful programmes, the creation of a set of decent provincial faculties of science.

Stanley Jaki's book is the first complete study of Duhem. This extensive treatment of the life and work of the author of the most famous book on theory in physics, *The Aim and Structure of Physical Theory* (1906), can only be welcomed with serious reservations. Its defects issue from Jaki's uncompromising hostility towards the Third Republic and most of Duhem's enemies. In addition to the distortions resulting from Jaki's gut Catholicism, there are basic misconceptions concerning the French educational system (including *grandes écoles* in the *Université*) and the history of science in France (that Darwinism was pushed in the faculties of science). Often the style seems to be modelled on the first sentence of *Areopagitica*. Still, this is a book on an "uneasy genius" that deserves to be chewed and digested. In the end, historical truth remains as elusive as scientific truth, which, as Houllevigue said, "resembles more a woman in a veil than the resplendent nude that painters make arise out of a well". □

H.W. Paul is Professor of History at the University of Florida, Gainesville, Florida 32611, USA. His Knowledge to Power: The Rise of the Science Empire in France, 1860-1939, will be published by Cambridge University Press later this year.

Beauty of defects

A.K. Jonscher

Defects and Defect Processes in Nonmetallic Solids. By W. Hayes and A.M. Stoneham. Wiley:1985. Pp.472. £54.30, \$47.

EARLY concepts of solid state physics were developed on the model of a perfect crystal — an ideal which proved extremely fruitful in generating a whole new branch of science. While it was recognized from the outset that defects were an inevitable and often a highly beneficial ingredient in any real solid, for a long time it was considered sufficient to regard them as small perturbations of the perfect host lattice and, in any case, to treat them as isolated and non-interacting entities. Only gradually did the impact of developing technology — of, for example, heavily doped and later amorphous semiconductors, or the progressive refinement of semiconductor devices or superionic conductors — demand a better understanding of the fundamental processes involving defects. The interaction of defects, exemplified by the phrase "ecology of solid state pollution", as a keyword-conscious researcher described simultaneous diffusion of two impurities in silicon, became increasingly a topic of detailed study.

Hayes and Stoneham's book is an important and probably a unique addition to the literature of this subject. It is remarkable in several respects. First, it combines outstanding readability with a depth of treatment which can be deceptive in that it can give the reader the impression that he has understood a certain point, until on closer examination he will find that on any single page there may be several other matters raised which require further study. The treatment is made both interesting and illuminating by the frequent inclusion of summary tables and other "integrating" features which bring together aspects of the subject not normally found under one heading, such as a full-page glossary of the main species of excitons.

The second unusual feature is the inclusion of many theoretical discussions which start with the simplest concepts and bring in, sometimes almost as asides, progressive refinements of approach which demand much wider reading. If one wishes to delve further, the up-to-date bibliography is very helpful; if not, one has at least been left under no illusion that the simple picture is the end of the matter.

The following chapter headings give an impression of the sheer breadth of the book's contents: "Electronic Properties" (mainly of perfect lattices but with electron-hole drops thrown in for good measure); "Interatomic Forces and Atomic Motions" (including defects and amorphous solids); "Lattice Defects" (with a discussion of fast ionic

conductors); "Spectroscopy of Solids" (with descriptions of optical, EPR and DLTS techniques); "Electronic Properties of Point Defects" (with a comprehensive review of defects in many different materials); "Radiation-induced Defect Processes" (ranging from radiation damage and ion implantation to laser annealing and photolysis); "Properties of Surfaces" (with special emphasis on defect phenomena and including a discussion of sintering); and finally "Special Systems" (which manages to include amorphous solids, metal-insulator transitions, intercalates and polymers).

The book will be a challenge to able final-year students — it will broaden their horizons but they should beware of becoming too engrossed in the detail — while for post-graduates this will be an ideal text to provide the background to many research topics. Teachers, too, will here find inspiration, though it may be necessary to dig deep into the subsidiary literature to follow up the items which are mentioned almost casually in the text. To a large degree, this is much more than a didactic text and has many features of a reference source. The seemingly effortless way in which the authors treat a wealth of topics should not deceive anyone into thinking that things are all that easy. □

A.K. Jonscher is Professor of Solid State Electronics in the Department of Physics, King's College London (KQC), Pulton Place, London SW6 5PR, UK.

NEW TITLES

THE ACID RAIN DEBATE SCIENTIFIC, ECONOMIC, AND POLITICAL DIMENSIONS

Edited by Ernest J. Yanarella and
Randal H. Ihara

This collection of essays by noted academics, lawyers, energy administrators, and research analysts focuses on the political and legal aspects of the acid rain debate, the policy options for resolving the controversy, and the international dimensions of acid rain control. The book complements the impressive body of research from the natural sciences and is supplemented by a comprehensive annotated bibliography on acid rain and relevant social science research.

Westview Press, September 1985,
c.200 pages, ISBN 0 8133 7065 5, pb, £18.00

ACID RAIN AND TRANSPORTED AIR POLLUTANTS IMPLICATIONS FOR PUBLIC POLICY

Office of Technology Assessment
Congress of the United States

Examining the subject from environmental, health, social, economic, and geographic viewpoints, this timely book gives a balanced coverage of all aspects of the acid rain controversy.

Unipub, 1985, 330 pages,
ISBN 0 89059 044 3, hb, £22.00

Stocked and distributed exclusively by

Bowker Publishing Co.
Erasmus House, 58-62 High Street,
Epping, Essex CM16 4BU
Tel: Epping 77333 (STD 0378)

Reader Service No.2

Problems, problems, problems . . .

Harry L. Shipman

Physics and Astrophysics: A Selection of Key Problems. By V. L. Ginzburg.
Pergamon: 1985. Pp.125. Hbk £16.75, \$23.50; pbk £8.95, \$12.50.

THE era of specialization is now well advanced. The last scientists to think of themselves as "physicists" rather than "solid-state physicists" or "high-energy physicists" are now nearing retirement age. Astrophysics is rapidly progressing in the same direction; some people who study galaxies have little appreciation of current research in stellar astronomy, and vice versa. One scientist once asked me, when I first met him, "What quarter acre of ground do you plough?". And much of the scientific literature, even review articles, consists of accounts of particular quarter-acres, with no attempt to describe the bigger picture.

Ginzburg's brief and entertaining book is a welcome exception to this trend and succeeds, for the most part, in describing what the entire landscape looks like, crossing the bounds of sub-specialities and giving a working scientist or student a feeling for what the person down the hall is really doing. He selects about 20 problems, dividing them into "macrophysics" (generally condensed matter, but including two problems in plasmas and one in nuclear physics), "microphysics" (particle physics) and astrophysics. You don't need to understand the specialized jargon in each of the fields in order to understand the book. (By specialized jargon I mean terms such as quantum Hall effect or Seyfert galaxy.) Each problem is discussed quite thoroughly, with a comprehensive set of references. Ginzburg has a remarkable, and welcome, international perspective — only a third of the hundreds of references are to literature in Russian, with the rest about evenly divided between American and non-American journals.

Physical intuition plays a large part in many of the chapters. Ginzburg leads the reader through back-of-the-envelope calculations, simple, easily understood derivations which tell you how, for example, the magnetization of some material depends on temperature, without considering whether fripperies such as the square root of π are part of the answer. It was only when I read an earlier edition of this book some years ago that I learned why my solid-state friends were going to great lengths to measure critical exponents — parameters that describe how important physical quantities change near a phase transition. A reader can understand not only what some important problems in fields far from his or her quarter acre of ground are, but why they fascinate the practitioners in that area.

Ginzburg has done an excellent job in keeping the book up to date, at least in including recent work on the problems

which he originally selected. The discovery of the W particle in 1983 is in here, as is the wait for the proton to decay, as are the ultrafast pulsars. I found only one place where he missed a beat — he introduces the spectroscopic and photometric measurements which indicated a concentration of fast-moving stars at the core of the giant elliptical galaxy M87, and which were interpreted as evidence for a black hole, but he doesn't mention subsequent work which showed that the evidence doesn't require a black hole.

Anyone who tries to select "key problems" — even in a subfield — has a lot of



V. L. Ginzburg — view beyond the back yard.

courage, and Ginzburg does more than just about anyone in taking on all of physics and astrophysics rather than just one tiny part. He makes no secret that "the subjective and controversial character of this book is quite apparent and readers have been warned — although, of course, such warnings are rarely heeded". He uses the problem of selecting "key problems", and the criticism he received as a result of it, as the springboard to several discussions of the way that science is done — just what it is that constitutes a "key problem", and the nature of scientific revolutions. He never precisely defines his selection criterion, but the book's contents indicate that he is interested mostly in unusual phenomena. Problems that appear include superconductivity, X-ray lasers, far-transuranic elements, grand unification theory, black holes and neutrino astronomy.

My reaction to the selection of problems is that, at least in astrophysics, Ginzburg's choice is a perfectly reasonable one, for the early 1970s. But the book shows its age. The first Russian edition appeared in 1971, and an English edition

was published in 1976. For this latest edition updating has been carried out primarily by appending text to each chapter. As a result, the discussion of these problems is quite current, but there are many hot topics which aren't in here at all. In condensed matter, I'll pick, rather arbitrarily, lower-dimensional materials and chaos as subjects of great contemporary interest which aren't included. Much of the excitement surrounding grand unified theories has been squeezed into a section originally entitled "CP invariance", but the cosmic connection between particle physics and cosmology is only alluded to briefly. The large-scale structure of the Universe, star formation, solar and stellar oscillations, the various evolutionary channels in late stages of stellar evolution, and dark matter, are other matters which are discussed only briefly, if at all.

Each reader, of course, will have his or her own list and it will undoubtedly include some problems which aren't in the book simply because those questions weren't asked when it was first written. Furthermore, the structure of the book dictates that some areas, such as biophysics, are not covered. Atomic and nuclear physics are also generally left out, and Ginzburg discusses much of the criticism which he has received for that choice, which was apparently deliberate.

Some of the problems covered are a bit dated. Testing general relativity, a key issue of the 1970s, has basically been completed in the weak-field limit (as Ginzburg himself points out), and it's at least less fashionable, if not less important, these days. Ginzburg makes a big issue of attacking (correctly, in my view) those astrophysicists who want to seek new physical laws around every corner, but once again the question of whether quasar redshifts are cosmological has subsided in the face of some overwhelming evidence that the redshifts of at least some, and probably all, quasars are produced by the expanding Universe.

There are some similar books around, which are the results of official commissions or review panels. But no matter how careful the deliberation of such panels, the impersonal style of any committee document generally makes it about as inspiring to read as an integral table. Ginzburg's book, however idiosyncratic and dated, is undoubtedly a good read; his imagination and insight come through on every page, and are well reflected in the English translation. Graduate students and young scientists should get hold of it, for it is one of the few places in print where one can acquire some idea of the exciting problems in contemporary physics and astrophysics. Those of us who have proceeded too far down the road of specialization will find in it a welcome view of the physical world beyond our own back yards. □

Harry L. Shipman is Professor of Physics at the University of Delaware, Newark, Delaware 19716, USA.

On minimizing the consequences of nuclear war

from Carl Sagan

The likely long-term global consequences of nuclear war remain catastrophic, despite attempts to minimize the dangers.

A FULL-SCALE nuclear war would be an unprecedented human catastrophe. But because we have no direct experience of such a war, the consequences and the policy implications must to some extent be uncertain. After an initial indication that soot from nuclear fires might provide significant atmospheric opacity¹, recent calculations (refs 2, 3 and subsequent investigations) have shown that the soot and dust attendant on nuclear war could, under many contingencies, cool and darken the Earth, with devastating global biological consequences⁴. This climatic catastrophe is called nuclear winter.

We believe our calculations^{2,3} to be a valid first-order assessment, although further research is needed to reduce uncertainties. We know of no factors that significantly and reliably reduce the possibility of grave worldwide climatic effects after a large-scale nuclear war. A survey by the US National Academy of Sciences⁵ concludes that there is a "clear possibility" both of "severe" consequences and of effects "that would be greater than, and would last longer than, that estimated in the baseline case". It is also possible that uncertainties might ameliorate the baseline climatic catastrophe. A study, by a committee appointed by the Royal Society of Canada⁶, concludes "we believe nuclear winter to be a formidable threat" and its analyses "tend to confirm that a drastic cooling will occur in the wake of a major nuclear war". An extensive analysis by a committee of the International Council of Scientific Unions^{4,9} concludes "There is a considerable probability a major nuclear war could gravely disrupt the global environment and world society."

Because these results and their implications for policy and doctrine⁶ are at considerable variance with the prevailing wisdom, it is natural and healthy that competent criticism be aired. However, in parallel with an inclination in the popular literature to exaggerate the consequences of nuclear war, there has also been a tendency in many professional pronouncements to down-play the consequences, a tendency which has been noted (and not

only with respect to nuclear war) in the risk-assessment literature^{4,5}.

In a recent *Nature* article⁷, in *The Reader's Digest*⁸ and in testimony before the Committee on Science and Technology of the US House of Representatives⁹, Edward Teller has demonstrated a dangerous propensity to minimize or ignore the consequences of nuclear war. He concludes that radioactive fallout, depletion of the protective ozone layer, and nuclear winter following a major nuclear war have all been "exaggerated". He artificially divides the total problem into small pieces, and then pronounces each piece "manageable".

Fallout

Teller concludes that fallout is "far less important than originally feared". Unaccountably he does not deal with the prompt fallout that would contaminate large areas of the target zones and adjacent territory in a nuclear war, eventually causing at least tens of millions of civilian fatalities¹⁰. In a major nuclear war, 30 per cent of Northern mid-latitude land areas, where the majority of people live, could receive fallout burdens approaching or exceeding the acute mean lethal dose for unprotected populations^{2,3}. By any standard, such effects would seem worthy of note.

On an intermediate timescale, still more widespread radioactive fallout would, over the years, cause further millions of cases of cancer and birth defects among the survivors of a nuclear war¹¹ and would produce serious effects synergistic with other consequences of nuclear war⁴. Teller says that the "best current calculations" from the Livermore weapons laboratory¹² indicate average long-term mid-latitude doses of about 20 rem, provided nuclear reactors, fuel reprocessing plants and spent fuel storage facilities are not targeted. But these calculations (still unpublished in a refereed journal) are based on relatively simple empirical scaling from nuclear test explosions, few of which were of a yield, or in a location, corresponding to a realistic war situation.

We independently^{2,3} used a microphysi-

cal model to predict likely radiation doses two to three times larger than did the Livermore study, with the same exclusion of nuclear power facilities. Neither model is entirely suited to this problem, but we must at least consider that the plausible range of such average long-term doses is 20 to 50 rem, at least an order of magnitude greater than the conventional wisdom, from which Teller did not, until recently, dissent. Both calculations also assume a fission yield fraction of 0.5, when smaller modern warheads may have fission yield fractions closer to 1.0.

Considering the other multiplicative factors quoted by Teller and by us^{2,3} (for 'hotspots' and concurrent internal doses), plus the computational factors mentioned, which increase the dose range by factors of 2.5 to 5, it seems clear that long-term radiation exposure could be much more serious than Teller assumes. The radiation dose would be still larger if more than 5,000 MT were detonated (more than twice this yield exists in the current strategic arsenals) or if nuclear fuel processing and storage sites were attacked, which is not unlikely in a major nuclear war.

When delivered acute whole-body doses exceed about 100 rem, the human immune system begins to be compromised; when coupled with other assaults on people and their environment occasioned by nuclear warfare, such doses can have extremely serious implications⁴. Nevertheless, prompt radioactive fallout, in almost every case, presents the more serious threat to mid-latitude populations. But fallout on an intermediate timescale could produce serious, unpredictable and continuing morbidity worldwide and for generations to come.

It is thus simply untrue that fallout is "far less important than originally feared". Fallout, not taken properly into account when nuclear weapons were first designed, began to be studied seriously only after the Bravo test of the Castle series at Bikini on 1 March 1954, which accidentally contaminated (with doses in excess of 100 rem) the atoll of Rongelap in the Marshall Islands, and the Japanese fishing

vessel *The Lucky Dragon*. Although some work on fallout was done as early as the Trinity Test in 1945¹⁴, not until Castle Bravo was there extensive public discussion of the issue, eventually contributing to the 1963 Limited Test Ban Treaty.

Ozonosphere depletion

On the effects of nuclear explosions on the ozone layer, it is misleading merely to compare the widespread systematic decreases in ozone caused by massive NO_x injections in a nuclear war with the natural fluctuations and latitudinal variations observed in the ambient ozone layer. Such fluctuations would be superimposed on a decrease by several tens of per cent caused by a large nuclear war.³ Here, as in other long-term consequences of nuclear war, it is important to distinguish between mean values and maximum excursions, both in time and in space. After a major nuclear war, average ultraviolet-B intensity would be increased above background by several factors for years, almost everywhere^{2,3}. Ecosystems are generally adapted to local ultraviolet fluxes, and serious damage can result from a sudden increase in exposure on this scale¹⁵.

Teller states that there are comparatively few nuclear weapons of yield sufficient to lift NO_x high enough to cause ozone depletion. At mid-latitudes, the fireball from a 100 KT explosion will just penetrate the lower stratosphere. Most strategic warheads deployed and planned by both nuclear alliances have warheads exceeding 100 KT and, especially when gravity bombs are included, span the rough yield range from 100 KT to 1 MT (with some older weapons ranging an order of magnitude higher and some submarine-launched ballistic missile warheads a factor ≈ 2 lower)^{16,17}. Teller overlooks the reinforcing effects of multiple nuclear explosions (such as from two or three bursts in rapid succession over one missile silo, with perhaps 100–200 MT detonated over a silo complex) in propelling nuclear debris to high altitudes, and the possibility of large-scale firestorm injection into the stratosphere. For these reasons, the present and future potential for ozonosphere depletion must continue to be monitored as an important secondary effect of nuclear weapons. It is noteworthy that the threat to the ozonosphere from nuclear weapons was first publicly identified not by those working on the development of nuclear weapons but by those studying an apparently unrelated problem, the effluents from proposed supersonic transport aircraft¹⁸.

Dust

Teller invokes a still classified Livermore report, with no designation number, on the climatic effects of dust. Perhaps the report alluded to is an early analysis of the dust mass in nuclear clouds from test

explosions, recently released by the cited authors as an informal and unclassified Livermore publication¹⁹ (and which does not mention climatic effects at all).

Teller told the US House of Representatives committee⁹ that, after assuming the directorship of the Lawrence Livermore National Laboratory in 1958, and after having been "troubled" by the possible climatic effects of dust "for decades before the publication of the ... report" (ref. 3), he initiated an early computer study: "Without that, ... all this discussion could not have taken place." In fact, none of the recent studies, and certainly none of our results, arise from these investigations, once again unpublished in refereed scientific journals. That during this period the US government was ignorant of the climatic consequences of nuclear war is made clear by the manner in which this topic is overlooked in the official study "*An assessment of frequently neglected effects of nuclear attacks*" prepared in 1978 and recently declassified²⁰. Moreover, there is no record of Teller having spoken publicly about such climatic effects before 1982.

Teller's assertions that "dust raised by a large-scale war is comparable with that produced in the largest volcanic eruptions" and that the "most probable conclusion is that the ... effects of dust would be noticeable but by no means severe on a hemispheric scale" have no quantitative foundation, and fly in the face of contrary evidence. Again, Teller divorces dust effects from all other simultaneous effects. Even isolated volcanic eruptions have been associated with striking weather anomalies, crop failures and significant human suffering. Through changes in the planetary albedo and general circulation, dust can have major climatic consequences^{2,3}.

Tests with chemical explosives are not a reliable basis for nuclear dust calculations. Unlike their chemical counterparts, nuclear explosions attain such great temperatures and pressures that vaporized and melted surface material contributes in a major way to the generation of fine dust particles (smaller than several micrometres in radius), and the steep thermal gradients can produce effects (such as 'popcorning') essentially unknown in chemical explosions. Tests of high explosives may, however, tell us something about the ejecta and sweep-up components of dust clouds, and place lower limits on these components in nuclear explosions. Multi-burst phenomena are also likely to augment dust production and injection heights over those expected from individual non-interacting explosions.

Smoke

Descriptions of severe climatic effects caused by smoke and dust in a nuclear war are based on detailed energy-balance

calculations, including a parametric treatment of important three-dimensional processes. In our calculations, we did not 'assume' large smoke emissions, but calculated them using data from the fire and civil defence literature. We selected the most probable values of key input parameters and considered likely variations of each parameter from that choice. Generally speaking, our climatic results, for almost fifty different cases, fall within a broad range of plausible outcomes that vary from significant to grave. The results remain within this range as key physical parameters describing dust and smoke production and their properties are varied over their intervals of uncertainty. In this sense, the results can be described as 'robust' (we do not insist on the term), although not 'conclusive' or 'proved'. Compared with our predictions, including our allowances for the thermal inertia of the oceans, the actual after-effects of a full-scale nuclear war are as likely to be more as less severe. Moreover, even small-scale nuclear wars may produce major climatic effects, on at least hemispheric scales, provided cities are targeted^{2,3}. Wind transport of dense smoke patches and streamers can produce "quick freezes" in distant localities very soon after the explosions²¹.

To mitigate nuclear winter, Teller seems to be invoking a meteorological miracle, involving various speculative 'abnormal mechanisms'. He mentions the importance of the water vapour budget of the atmosphere. But we explicitly accounted for water cycling and scavenging by precipitation in our calculations of smoke lifetime, employing the conservative assumption that smoke is efficiently scavenged, which may not be true. Up to 50 per cent of all smoke was taken to be removed in immediate black rain. There is therefore no evidence that the water issue raised by Teller is more important than already allowed for in our calculations.

Capping clouds. One of Teller's arguments is that capping clouds will form above large fires, accelerating smoke removal. But such clouds do not occur over all fires; their occurrence depends on the local meteorology. It is not even clear that smoke particles contained in a capping cloud are removed from the atmosphere; micrometre-sized cloud particles are different from millimetre-sized precipitation particles. L. Radke observed a capping cloud over a prescribed forest burn in 1978⁵, and by good chance measured the smoke properties in two dissimilar air parcels within the plume: one parcel had passed through the capping cloud, the other had not. In the cloud-processed smoke, the concentration of the smallest particles, $\leq 0.05 \mu\text{m}$ radius, was less by a factor of about 10 than in the free-drifting smoke, but the total smoke mass was comparable in both parcels. The larger

particles were similar in number, and the measured optical scattering coefficients were essentially indistinguishable. In other words the processed and unprocessed smokes would have nearly equal climatic impact. (The Stokes-Cunningham fall velocities of these smoke particles are entirely negligible.) Downwind of wild-fires, the submicrometre smoke particle size distributions are usually very similar^{22,23}, which further suggests that cloud processing is not significant in the average case.

Washout in storms. We made separate climate calculations for ocean and land environments, using knowledge of ocean/continent wind patterns and the normal climate to estimate the ameliorating effects of the oceans on land temperatures^{2,3}. Thus, while we did not perform a three-dimensional calculation, the mean effects of oceans were approximately included "in a quantitative manner". Our estimates were later supported by Livermore²⁴, NCAR^{21,26,47}, Los Alamos⁴⁸ and Soviet^{25,26} two- and three-dimensional calculations for similar conditions. (The important Livermore calculations unfortunately remain unpublished in refereed journals.) The ameliorating effect of oceanic temperature must, of course, be limited; otherwise natural winters would not occur every year. While storms at ocean/continent margins would remove some of the lower-level smoke, they would also carry some to higher altitudes. Further work is needed on initial and subsequent removal rates and on the height of the stabilized smoke cloud.

Teller's speculation that smoke removal might be self-accelerating is unsound. Satellite photographs of fire plumes²⁷ typically show smooth and rapid dispersion over many hundreds of kilometres, with no indication of induced turbulence and washout. Furthermore, enhanced spreading and stabilization of dark oil fire plumes have been observed²⁸.

Smoke nucleation. The nucleating ability of smoke particles depends on their origin. In forest fires, where induced winds are expected to lift quantities of ash, char and soil debris, a substantial fraction of the particles (around 10 per cent) may act as water condensation nuclei²⁹. However, in an oil fire plume without extraneous contaminating particle sources, there appears to be no significant increase of nuclei concentration over background³⁰. There is also substantial observational evidence that fire plumes do not enhance, and may indeed suppress, rain formation in cane field fires³¹, forest fires²⁹ and oil fires³⁰.

The origins of the black rains at Hiroshima and Nagasaki have not been unambiguously determined, although the phenomenon is clearly related to the large fires. Recent fire plume simulations based on detailed cloud physics models³² suggest

that smoke rainout may not be as frequent or as efficient as has been inferred from the anecdotal accounts at Hiroshima, or as was assumed in our calculations. A range of recent studies, presented at a Defense Nuclear Agency Conference on Large-Scale Fire Phenomenology (National Bureau of Standards, 10-13 September 1984), suggests that intense urban firestorms can propel tens of per cent of fine soot particles to high altitudes. (See also ref. 32.) Solar radiative heating of low-altitude soot can also propel it to even higher altitudes, where the residence time is approximately one year, and the coverage can rapidly become global.^{3,48} These are additional respects in which the calculations in ref. 3 may be conservative.

Teller's mechanism for smoke removal depends on at least two unsupported contentions:

- (1) that turbulent coastal weather can wash out a significant fraction of nuclear smoke before it can be widely dispersed; and
- (2) that smoke clouds tend to become more patchy, not less, as they disperse through the atmosphere, which is contradicted by direct observations of pollutant tracers.

Urban smoke estimates. Broyles³³, the Livermore consultant whose estimates of urban smoke emission are cited by Teller, relies heavily on forest fire data, and does not review the urban fire literature or discuss smoke emissions characteristic of urban materials and conflagrations. Thus "proper modifications" for the urban case have not all been made. For example, Broyles, in his analysis, adopts smoke emission factors ranging from 0.1 per cent to 6.3 per cent, but the lower value is clearly inapplicable to cities, with their fossil fuel stocks and plastics. Accordingly, it is likely that Broyles' lower limit to smoke emission of 15 Tg (incidentally, not a strict limit) greatly understates the problem. However, even 10 Tg of sooty smoke, uniformly distributed, could decrease the average amount of sunlight reaching the Earth by one-third. Moreover, because of the necessarily uneven burning and restricted ventilation of large fires, more smoke than expected may be produced; experimental data show that restricted air flow greatly increases smoke generation³⁴.

Assessments of the quantities of "fuels" available in urban areas have been made by several groups^{3,35,36}. These assessments are all in general agreement, although more detailed information is clearly needed. Teller notes that 150 Tg of "fuel" is available at US oil refineries³⁷. However—including refineries throughout North America, the Soviet Union, Europe and the Middle East, as well as fuels in storage at transit nodes in urban areas, processed flammables in tankage, warehouses and transport, and such end

products in use as roofing, paving and plastics—the actual figure for available fossil fuels and products seems to be ten times Teller's figure, or some 1,500 Tg (refs 5, 36). US strategic oil reserves in underground salt domes amount to about 80 Tg (the reserves are currently about one-half full). In some cases, the oil is close enough to the surface to pool and burn within a weapons-generated crater³⁸.

While Teller agrees that 10 per cent of the fossil fuel compounds would be converted to smoke when burned, some 60 per cent or more of the smoke would consist of soot with large absorption coefficients in the visible, not the 10 per cent he assumes^{39,40}. Hence, Teller's estimate of potential soot emission from fossil-based fuels alone is low by a factor of some $10 \times 6 = 60$. A total petroleum-derived soot emission of 90 Tg is possible if major urban/industrial areas are attacked—sufficient to reduce average direct global sunlight intensities to 4% of ambient.

Realistically, absorbing and scattering particles from sources other than petroleum should also be included in the calculations. Urban combustibles (for example wood, paper, cloth, paint, household chemicals, rubber and synthetic fibres) could, we calculate, release another 10 to 100 Tg of smoke. Wildfires ignited by nuclear explosions over forest and grasslands could emit anywhere from 0 to 50 Tg of smoke. Dust would also be generated by surface nuclear detonations; attacks on hard military targets might add 10 to 100 Tg of fine scattering particles to the upper atmosphere. Powerful convective winds over fires could raise urban debris (for example, masonry dust, metal smoke, fine ash) in large quantities. In summary, it does not seem difficult, in a wide range of nuclear war scenarios, to loft sufficient sooty smoke and dust to initiate major climatic effects—especially if cities are burned.

It is also worth mentioning that explosions and fires in urban/industrial areas would release, suddenly, large quantities of toxic chemicals (pyrotoxins) into the atmosphere^{2,3}. This problem remains to be fully quantified, and could provide the next departure from the conventional wisdom in studies of the consequences of nuclear war.

In his Congressional testimony, Teller⁹ asserted "If you reduce the smoke by a factor of two, then instead of a temperature decrease which is sizeable, you get temperature changes which are in some cases negative, in some cases positive... The calculations so far... seem to show that about one-half of the smoke assumed in the paper of Turco and others, about one-half the smoke is already below the threshold [for significant climatic effects]. Now if somebody is dedicated to exaggeration, we can't stop him." In fact,

as long as the smoke is injected into the upper troposphere, the abundance can be one-tenth that in our baseline case³, or around 20 Tg, and serious climatic consequences, of at least hemispheric scale, will still follow. Regional-scale effects are likely with even smaller soot loadings.

New weapons

Teller claims that, because of a trend towards more accurate lower-yield weapons, collateral damage, ozonosphere depletion, fires and nuclear winter will, in the future, be much less likely to occur; and also that ballistic missile defence would further reduce the threat of fires and smoke. But most new strategic weapons deployments in the next decade or two will apparently involve warheads in the 100 KT to 1 MT range^{16,17}. A 100 KT airburst can destroy up to 50 km². Weapons of this yield, used against industrial areas, are the ideal incendiary devices for adjacent or co-located cities; they have a quick thermal pulse, and do not produce large areas of deep rubble. Thus, the reconfiguration of nuclear arsenals now underway may be, inadvertently, a move towards a more effective stockpile of nuclear weapons for burning cities and triggering nuclear winter—unless it is accompanied by a dramatic reduction in the strategic arsenals. In addition, the Minuteman III and SS-18 missiles, with their 300 to 600 KT warheads, are likely to be with us for some time to come.

It is extremely shortsighted to assume

that defensive measures against ballistic missiles will not stimulate new offensive deployments. Effective defence systems are not technically feasible or even designed, could not be deployed for decades, and when deployed would be vulnerable to direct attack, to underflying (for example, by aircraft and cruise missiles) and to relatively simple counter-measures⁴¹⁻⁴⁴. The cities—and therefore the entire Northern Hemisphere, at least—will remain in serious jeopardy for the foreseeable future. It is not apparent that the peril can be significantly reduced without a major reversal of the nuclear arms race. Other reasons that low-yield weapons and strategic defence are unlikely to resolve the issues raised by nuclear winter are discussed elsewhere⁶.

Teller's conclusions

Here, while Teller reiterates some of his arguments, he concludes, nevertheless, that severe climatic perturbations *could* occur after a nuclear war, threatening world populations. The only solution that he puts forward is to store food for survivors in the United States and some other (unspecified) nations.

Teller writes, "Highly speculative theories of world-wide destruction—even the end of life on Earth—used as a call for a particular kind of political action serve neither the good reputation of science nor dispassionate political thought." But, before it was widely known that there were dangers from fallout, or

from ozonosphere depletion, or from climatic effects, we were not being warned that the alleged *absence* of widespread or long-term effects was speculative, or that the consequences of nuclear war might be more severe than asserted by those responsible for national policy and many of their scientific advisers. However, when a range of scientific evidence suggests that these effects are real, with grave biological and social consequences, then Teller comes forth to minimize the possible consequences, reminding us that not all the evidence is in. This, I suggest, is a clear double standard of scientific evidence at work. Teller's attempt to minimize the widespread and long-term consequences of nuclear war has in fact made possible "a particular kind of political action"—namely, the accession of ever-increasing numbers of weapons of mass destruction, beyond any military utility, and far beyond the requirements of strategic deterrence.

I have benefited enormously from discussion with my co-authors in refs 2 and 3, R. P. Turco, O. B. Toon, T. P. Ackerman, and J. B. Pollack; I am grateful to them, especially R. P. Turco, and to Hans Bethe, Paul Crutzen, Ann Druyan, Richard L. Garwin, Lester Grinspoon, John Holdren, Dorothy Nelkin, Stephen H. Schneider and Steven Soter for a critical reading of this paper.

Carl Sagan is at the Laboratory for Planetary Sciences, Cornell University, Ithaca, New York 14853-0355, USA.

- Crutzen, P. & Birks, J. *Ambio* 11, 114-125 (1982).
- Turco, R. P., Toon, O. B., Pollack, J. B. & Sagan, C. *EOS* 63, 1018 (1982).
- Turco, R. P., Toon, O. B., Ackerman, T. P., Pollack, J. B. & Sagan, C. *Science* 222, 1283-1292 (1983). ["TTAPS"].
- Ehrlich, P. R. *et al. Science* 222, 1293-1300 (1983).
- Carrier, G. *et al. The Effects on the Atmosphere of a Major Nuclear Exchange* (National Academy of Sciences, Washington, DC, 1985).
- Sagan, C. *Foreign Affairs* 62(2), 257-292 (Winter 1983/84).
- Teller, E. *Nature* 310, 621-624 (1984).
- Teller, E. *Reader's Digest*, 138-144 (November 1982).
- Teller, E. *Statement before the Committee on Science and Technology, Subcommittee on Natural Resources, Agricultural Research & Environment, U.S. House of Representatives* (12 September 1984).
- The Effects of Nuclear War*, Office of Technology Assessment, OTA-NS-89 (May 1979).
- Ambio* 11, 2-3 (1982).
- Knox, J. B. *Global Scale Deposition of Radioactivity from a Large-Scale Exchange* (Publ. UCRL-89917, Lawrence Livermore National Laboratory, 1983).
- Chester, C. V. & Chester, R. O. *Nucl. Tech.* 31, 326-338 (1976).
- Badash, L., Hirschfelder, J. O. & Broda, H. P. *Reminiscences of Los Alamos, 1943-1945*, 73-78 (Reidel, Dordrecht, 1980).
- Causes and Effects of Changes in Stratospheric Ozone: Update 1983* (National Academy of Sciences, Washington, DC, 1984).
- Arkin, W. M., Cochran, T. B. & Hoenig, M. M. *Bull. Atomic Sci., Suppl.* August/September (1984).
- Cochran, T. B., Arkin, W. M. & Hoenig, M. M. *Nuclear Weapons Databook: Vol. I, U.S. Nuclear Forces and Capabilities* (Ballinger, Cambridge, Massachusetts, 1984).
- Foley, H. M. & Ruderman, M. A. *J. geophys. Res.* 78, 4441-4450 (1973).
- Gutmacher, R. G., Higgins, G. H. & Tewes, H. A. *Total Mass and Concentration of Particles in Dust Clouds* (Publ. UCRL-14397, Lawrence Livermore National Laboratory, 1983).
- An Assessment of Frequently Neglected Effects of Nuclear Attacks, U.S. Arms Control & Disarmament Agency, Civil Defense Study No. 5*, April 19 (1978).
- Covey, C., Schneider, S. H. & Thompson, S. L. *Nature* 308, 21-25 (1984).
- Sandberg, D. V., Pierovich, J. M., Fox, D. G. & Ross, E. W. *US Forest Service Tech. Rep. WO-9* (1979).
- Ward, D. F., Nelson, R. M. Jr & Adams, D. F. *Forest Fire Smoke Plume Documentation* (US Department of Agriculture, Macon, Georgia, 1979).
- MacCracken, M. C. *Nuclear War: Preliminary Estimates of the Climatic Effects of a Nuclear Exchange* (Publ. UCRL-89770, Lawrence Livermore National Laboratory, 1983).
- Alexandrov, V. V. & Stenchikov, G. L. *On Modelling the Climatic Consequences of Nuclear War*, Proc. appl. Math. (Computer Center, USSR Academy of Sciences, 1983).
- Thompson, S. L. *et al. Ambio* 13, 236-243 (1984).
- Matson, M., Schneider, S. R., Aldridge, B. & Satchwell, B. *NOAA Tech. Rept. NESDIS 7*, (US Department of Commerce, Washington, DC, 1984).
- Davies, R. W. in *Atmospheric Diffusion and Air Pollution* (eds Frenkel, F. N. & Sheppard, P. A.) 413-415 (Academic, New York, 1959).
- Eagan, R. C., Hobbs, P. V. & Radke, L. F. *J. appl. Meteorol.* 13, 553-557 (1974).
- Radke, L. F. *et al. Proc. 3rd WMO sci. Conf. Weather Modification*, Clermont-Ferrand, France, July 21-25, 119-126 (1980).
- Mordy, W. A. *Geophys. Monogr.* 5, 399 (American Geophysical Union, 1960).
- Cotton, W. R. *Am. Sci.* 73(3), 275 (1985).
- Broyles, A. A. *Am. J. Phys.* 53(4), 323 (1985).
- Tewarson, A. & Steciak, J. "Fire ventilation", Factory Mutual Research, Norwood, Mass., Technical Report FMRC J.I., OEON 6C 070(a), October (1982).
- FEMA Attack Environment Manual*, CPG 2-1A3, (Federal Emergency Management Agency, Washington, DC, 1982).
- Crutzen, P. J., Bruhl, C. & Galbally, I. E. *Climat. Change* 6, 323-364 (1984).
- Oil Gas J.* 82, 114 (1984).
- International Petroleum Encyclopedia*, 316-322 (Pennwell, Tulsa, 1984).
- Day, T., MacKay, D., Nadeu, S. & Thurier, R. *Wat. Air Soil Pollut.* 11, 139-152 (1979).
- Tewarson, A. in *Flame Retardant Polymeric Material Vol. 3* (eds Lewin, M., Atlas, S. M. & Pierce, E. M.) 97-153 (Plenum, New York, 1982).
- Bethe, H. A., Garwin, R. L., Gottfried, K. & Kendall, H. W. *Scient. Am.* 251, 39 (1984).
- Tirman, J. (ed.) *The Fallacy of Star Wars* (Random House, New York, 1984).
- Carter, Ashton B. *Directed Energy Missile Defense in Space*, Office of Technology Assessment, U.S. Congress (1984).
- Drell, S. D., Farley, P. J. & Holloway, D. *The Reagan Strategic Defense Initiative: A Technical, Political and Arms Control Assessment* (International Strategic Institute, Stanford University, 1984).
- Slovic, P., Fischhoff, B. & Lichtenstein, S. in *Societal Risk Assessment* (Plenum, New York, 1980).
- Hare, F. K. *et al. Nuclear Winter and Associated Effects* (Royal Society of Canada, 1985).
- Thompson, S. L. *Nature* (submitted).
- Malone, R. C., Auer, L. H., Glatzmaier, G. A., Wood, M. C. & Toon, O. B. *J. geophys. Res.* (submitted).
- Scientific Committee on Problems of the Environment, International Council of Scientific Unions, *Environmental Consequences of Nuclear War* (Wiley, New York, in the press).

The hepatitis B virus

Pierre Tiollais, Christine Pourcel & Anne Dejean

Unité de Recombinaison et Expression Génétique, INSERM U163, CNRS LA 271, Institut Pasteur, 28 rue du Dr Roux, 75015 Paris, France

DNA recombinant technology has radically changed hepatitis B virus (HBV) virology. The genetic organization, transcription and replication of the virus are basically understood, structures of integrated HBV sequences in hepatocellular carcinoma have been characterized, and new vaccines produced by recombinant DNA technique are being developed.

HEPATITIS B is a public health problem of worldwide importance. In Far-East Asia and tropical Africa, chronic carriers of the hepatitis B virus (HBV) represent 10% or more of the population and chronic active hepatitis and liver cirrhosis are major causes of mortality. Moreover, epidemiological studies have clearly shown the importance of HBV in hepatocellular carcinoma (HCC), one of the most common cancers in the world¹. In China, between 500,000 and one million new cases appear every year. HBV is one of the few viruses known to be involved in a human cancer.

HBV infection is highly polymorphic, ranging from inapparent forms to acute hepatitis and severe chronic liver disease. The pathological consequences of the viral infection are unpredictable and the mechanism of liver damage is not fully understood. The virus is not cytolytic and the host immune response to viral antigens present at the liver cell membrane has a major role in the pathogenesis of HBV-related liver diseases².

HBV belongs to the group of animal viruses known as the hepadna viridae³. The other members of this group are the woodchuck hepatitis virus (WHV)⁴, the beechey ground squirrel hepatitis virus (GSHV)⁵ and the Pekin duck hepatitis B virus (DHBV)⁶. All these viruses have a common structure. They are mainly hepatotropic and lead to persistent virus infection. Only HBV and WHV cause chronic active hepatitis and HCC.

Hepatitis B virus particles

During HBV infection in humans, virus particles are present in very large quantities in the blood (Fig. 1). In some forms of chronic infection, the serum contains only empty viral envelopes. Occasionally, it also contains complete virions, the empty envelopes always remaining in large excess. The presence of complete virions in the serum is indicative of an active viral multiplication in the liver. The empty envelopes consist of spherical or filamentous particles 22 nm in diameter. The virion of 42 nm diameter consists of an envelope and a nucleocapsid containing a circular DNA molecule, a DNA polymerase, a protein kinase activity and a 'DNA-linked protein'. The envelope carries the hepatitis B surface antigen (HBsAg) and the capsid, the hepatitis B core antigen (HBcAg). When virions are present in the blood, an additional soluble antigen related to the capsid, the hepatitis B e antigen (HBeAg), is generally detected in the serum. HBsAg contains a group determinant called a and subtype determinants termed d, y, w and r. Three main serotypes, adw, adr and ayw, are commonly observed but have an unequal distribution worldwide⁷. The y and r determinants are, respectively, absent from Far-East Asia and from Africa.

The envelope of HBV consists of proteins, carbohydrates and lipids. Glycoproteins are anchored in a lipid bilayer. Electrophoretic analysis and comparison of tryptic digests with the sequence of the envelope genes recently led Gerlich and his colleagues to redefine the protein composition of the envelope^{8,9}. The envelope of the virion contains three proteins called the major, the middle and the large proteins. The major protein is 226 amino acids long and is encoded by the S gene (Fig. 2). It exists in two forms, glycosylated (GP27^S) and non-glycosylated (P24^S). GP27^S possesses a complex N-linked glycan at Asn 146.

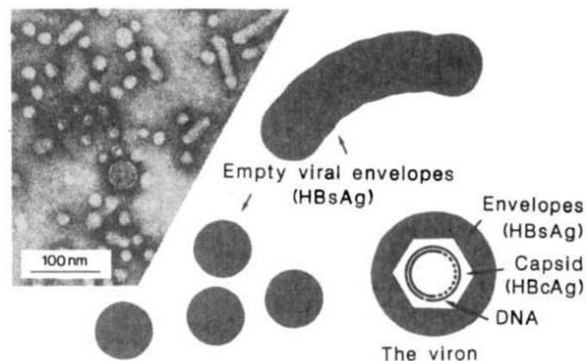


Fig. 1 The hepatitis B virus particles. Left: An electron micrograph of the viral particles present in the serum of an HBV chronic carrier (micrograph courtesy of A. J. Zuckerman). Numerous 22-nm spherical particles and filaments and one virion (or Dane particle) are present. Right: A schematic representation of the viral particles. The virion consists of an envelope carrying HBsAg and a capsid carrying HBcAg and containing the circular DNA molecule. The 22-nm spherical particles and the filaments are empty envelopes carrying HBsAg.

Analysis of its primary structure deduced from the nucleotide sequence shows that it contains three hydrophobic sequences separated by two hydrophilic ones. Precise limits of these segments cannot be defined, but for the three hydrophobic segments they could be amino acids 7–23, 80–98 and 169–226¹⁰. The two first hydrophobic sequences probably traverse the lipid bilayer. Most of the amino-acid variations observed in the different sequenced HBV genomes occur in the two hydrophilic domains.

HBsAg is a conformational antigen. A dimer of two major protein molecules linked by disulphide bridges represents the structural unit that bears full HBsAg antigenicity¹¹. Reduction of disulphide bonds results in dissociation of the dimer and a drastic decrease of HBsAg antigenicity. The segment of the major protein between residues 122 and 155, in the second hydrophilic domain, is an exposed¹² region and results obtained using synthetic oligopeptides have shown that this domain contains most of the HBsAg group and subtype determinants^{13–18}.

The middle protein is 281 amino acids long and is encoded by the pre-S2 region and S gene (Fig. 2)⁸. It is a glycoprotein present in two forms, GP33^S and GP36^S, according to the extent of glycosylation; GP33^S contains one glycan and GP36^S two glycans. The 55 amino acids encoded by the pre-S2 region are in general hydrophilic and contain a dominant epitope located at the surface of the envelope^{19,20}. This epitope is disulphide bond-independent and apparently more immunogenic than are epitopes of the major protein. The pre-S2-encoded sequence contains a receptor for polymerized human serum albumin (pHSA)²¹. The hepatocytes also have a receptor for pHSA and it has been postulated that pHSA mediates the attachment of HBV to hepatocytes.

The large protein is encoded by the pre-S1 region, pre-S2 region and S gene (Fig. 2) and is present in a glycosylated (GP42^S) and a non-glycosylated (P39^S) form⁹. This protein is

of variable length according to the subtype, 389 or 400 amino acids long for the ay and ad subtypes, respectively. GP42^S contains one N-linked glycan. At least part of the N-terminal sequence, encoded by the pre-S1 region, is located at the surface of the envelope. The pre-S1 translation product may also be involved in the attachment of HBV to hepatocytes²⁰.

One virion contains 300–400 major protein molecules and 40–80 middle and large protein molecules⁹. The protein composition of filaments is identical to that of the complete virions while the composition of the 22-nm diameter empty spherical particles is quite different. In the serum of chronic carriers with viral replication, the spherical particles contain both the major and middle proteins in about the same ratio as in the virions but with at least 20 times less of the large protein. This may be the result of a difference between the mechanism of assembly of virions and of empty envelopes. In the absence of viral replication, the spherical particles contain mostly the major protein, and <1% of middle protein and no large protein. The relative rates of synthesis of the three envelope proteins could be regulated at the transcriptional level.

The nucleocapsid of HBV consists of one major protein, the core protein (P22^C), and contains a protein kinase activity capable of phosphorylating the P22^C protein^{22,23}. The C-terminus of P22^C is remarkable in that it is extremely rich in arginine, serine and proline residues²⁴. Presumably, the arginine-rich sequences, in which there is a tandem repeat of the octapeptide Ser-Pro-Arg-Arg-Arg-Ser-Gln, are involved in interactions with the HBV genome within the nucleocapsid. Such sequences resemble protamine and other DNA binding proteins. HBcAg is composed of both conformational and linear determinants of the P22^C protein. HBcAg is a cryptic determinant released by disruption of the capsid²⁵, and is present in the serum as a protein of 15,000 (15 K) relative molecular mass (M_r) which results from proteolytic cleavage of P22^C. The C-terminal sequence Thr-Thr-Val-Val was determined for serum HBcAg. This sequence lies 33 amino acids from the end of P22^C, within the sequence Thr-Thr-Val-Val-Arg-Arg²⁶. The HBV genome has an unusual and characteristic structure (Fig. 2)²⁴. It is a small circular, partly double-stranded DNA molecule with a single-stranded DNA region of variable length. The long or L(–) strand is linear and of a fixed length of about 3,200 nucleotides. The short or S(+) strand is of variable length, ranging from 50 to 100% of that of the L(–) strand. The positions of the 5' ends of the L(–) and S(+) strands are fixed, while the position of the 3' end of the S(+) strand is variable. The maintenance of the circular structure of the genome is assured by base-pairing of the 5' ends of the two strands. The 5' ends of the S(+) and L(–) strands were recently mapped to positions 1,601 and 1,826 (H. Will, personal communication). Interestingly, at both sides of the cohesive ends there is an 11-base-pair (bp) direct repeat (DR) 5'TTCACCTCTGC (Fig. 2). The two copies of this sequence, which start at nucleotides 1,824 and 1,590, are termed DR1 and DR2, respectively. That a similar direct repeat is found in all the hepadna viruses argues for a biological role for the DR sequence (Fig. 3).

Genetic organization of HBV

The restricted host range of HBV, which infects only man and chimpanzee, and the lack of a cell culture system for its propagation have greatly impeded our understanding of its molecular genetics. A general presentation of the genetic organization of the hepadna viridae was deduced from the comparative analysis of the nucleotide sequences of cloned genomes²⁴. Large open reading frames (ORFs) conserved among all the sequenced genomes probably represent coding regions for viral proteins (Fig. 2).

The L(–) strand transcript of HBV contains four ORFs conserved in the sequences of different strains. Moreover, insertions or deletions present in HBV subtypes are always multiples of three, allowing conservation of the reading frame^{27–32}. In contrast, there are no conserved ORFs of any considerable length in the S(+) strand transcript. The L(–) strand thus carries

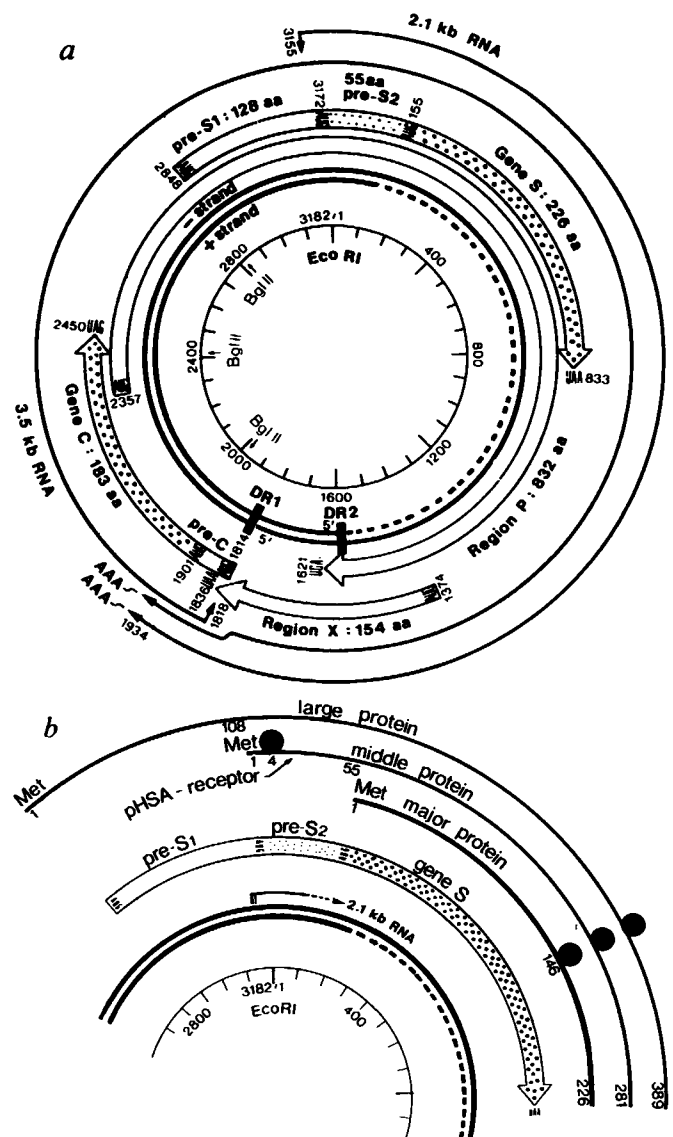


Fig. 2 Structure and genetic organization of the HBV genome. The broad arrows surrounding the genome represent the four large open reading frames of the L(–) strand transcript. The number of amino acids (aa) encoded by the coding sequences are indicated. The two thin arrows surrounding the broad arrows represent the two major HBV mRNAs. The partial restriction map and the numbering of the nucleotides indicated on the inner circle correspond to the ayw₃ genome²⁷. DR1 and DR2 are, respectively, at positions 1,824 and 1,590. The positions of initiation and stop codons of the coding sequences and the 5' and 3' ends of the two major transcripts are indicated.

virtually all of the protein coding capacity of the virus and can therefore be considered as the minus strand. The four ORFs of the HBV L(–) transcript are termed S, C, P and X. The P region overlaps with all the other three. The L(–) strand in this presentation is read one-and-a-half times¹⁰.

The S region codes for the protein of the viral envelope and is divided into the S gene, pre-S1 region and pre-S2 region (Fig. 2). The S gene and pre-S2 region are of constant length in all the HBV subtypes. The 5' end of the pre-S1 region of ad subtypes contains an additional sequence of 33 or 12 bp compared with ay subtypes³³. The heptanucleotide sequence 5'GCTAGGG directly repeated at both extremities of the additional sequence could be related to its insertion or deletion. In general, there is about twice as much amino-acid variability in the pre-S1 and pre-S2 regions as in the S gene. This suggests that the amino-acid sequence of the products of the pre-S regions are less critical for virus assembly than that of the S gene. Variations of the

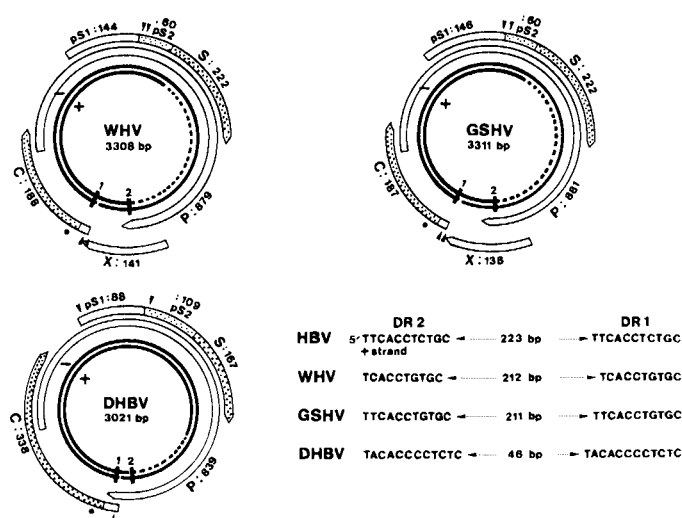


Fig. 3 Genetic organization of the hepadna viruses. The broad arrows surrounding the genomes represent the large open reading frames of the L(-) strand transcript. The numbers indicate the lengths of the coding sequences. The short arrows perpendicular to the broad arrows indicate the initiation sites of transcription of the major viral mRNAs. The asterisks indicate the common position of the 3' end of the viral mRNAs. DR1 and DR2 are respectively indicated as 1 and 2. The sequence of the DR of the four hepadna viruses and the distance between DR1 and DR2 are represented at bottom right.

pre-S regions could be a means of evading the host immune system.

The C gene encodes the core protein. The N-terminal amino-acid sequence of P22^C is unknown and two AUG at positions 1,814 and 1,901 (Fig. 2) are candidates for the initiation of translation of P22^C. According to the mapping of the 5' end of the P22^C messenger RNA (H. Will, personal communication), the AUG at position 1,901 should be considered as the initiation codon, and we will refer to the region between AUG 1,814 and 1,901 as the pre-C region. The coding capacity of the pre-C has been shown *in vitro* using expression vectors. Initiation of translation at the pre-C initiation codon produces a hydrophobic polypeptide bearing HBeAg and which could have a role in the attachment of core to the viral envelope (ref. 10 and W. H. Gerlich and W. Rutter, personal communications).

The translation product of the P region is a basic protein, rich in histidine, of ~90,000 M_r, the relative molecular mass of DNA polymerases. An amino-acid sequence comparison between the translation product of the P region and the reverse transcriptase (*pol* gene product) of all retroviruses sequenced to date and of cauliflower mosaic virus (CaMV) reveals tracts of homology in a region that overlaps the end of the S gene³⁴. The P region, therefore, certainly encodes the viral DNA polymerase which possesses a reverse transcriptase activity.

The X region can code for a polypeptide of 145–154 amino acids depending on the subtype. The function of this region is unknown, but recently Moriarty *et al.*³⁵ have detected in HBV-infected liver a polypeptide (P28) that reacts with antibodies against synthetic peptides corresponding to the X region. Antibodies against the translation product of this region were also detected in sera of patients with HCC^{35,36}.

The overall structure of the genomes of the hepadna viruses is similar (Fig. 3). The size of the WHV and GSHV genomes is around 3,300 bp^{37,38} and that of DHBV 3,000 bp³⁹. Nucleotide homologies between HBV and the three animal hepadna viruses WHV, GSHV and DHBV are around 70, 55 and 40%, respectively. WHV and GSHV sequences are very closely related, with >80% homology. The genetic organization of WHV and GSHV are also very similar to each other and to that of HBV. Only the pre-S1 nucleotide sequences of WHV and GSHV differ completely from that of HBV because of the presence of large insertions and deletions. The genetic organization of DHBV

differs somewhat from those of the mammalian hepadna viruses. The L(-) strand transcript contains only three ORFs: S, C and P. The 5' end of the DHBV C gene shows no homology with the X gene of mammalian hepadna viruses. Therefore, the DHBV C gene cannot be considered as a fusion sequence between gene C and region X. The absence of the X region in DHBV raises the question of its possible role in the other hepadna viruses.

The nucleotide sequences and genetic organization of mammalian hepadna viruses show close homology. Nevertheless, the pathology of GSHV, characterized by an asymptomatic persistent infection and no HCC, is closer to that of DHBV than to other mammalian hepadna viruses. Variations in the host immune response to the different hepadna viruses may explain these differences. Construction of recombinants between WHV and GSHV and subsequent infection of either woodchuck or ground squirrel could help to elucidate the role of specific viral sequences in the pathology, including HCC, of these two viruses (C. Seeger and H. Varmus, personal communication). Moreover, the absence of chronic hepatitis and HCC in chimpanzee infected with HBV indicates the importance of the host response in the liver diseases.

Transcription of HBV

In the absence of cell culture systems to propagate HBV, chronically infected liver of chimpanzee constitutes the most suitable model for studying viral transcription during virus replication. Permanent cell lines established from human HCC containing integrated HBV sequences and transfected cell lines with cloned HBV DNA are also useful for studying HBV transcription in the absence of viral replication. This last system allows the precise control of the HBV DNA fragments and especially of the regulatory elements introduced into the cell.

Two major HBV-specific poly(A)⁺ RNAs were characterized in chimpanzee infected liver^{40,41}. Both are L(-) strand transcripts and are referred to as the 2.1-kilobase (kb) and the 3.5-kb RNA (Fig. 2). The 5' end of the 2.1-kb RNA was mapped to ~20 bp upstream of the pre-S2 region and the 3' end to the very beginning of the C gene. It covers the pre-S2 region, gene S and region X. The 5' end of the 3.5-kb RNA was mapped to the pre-C region (H. Will, personal communication) and the 3' end to the same position as that of the 2.1-kb RNA. It covers the complete genome plus an additional sequence of ~100 nucleotides. This implies that the 3.5-kb RNA is produced from covalently closed double-stranded DNA and processed only at the second run of transcription. Minor RNA species of larger and smaller sizes were also detected. No splicing event seems to exist in HBV transcripts, at least concerning the two major RNAs. Cell transfection experiments with cloned HBV DNA fragments and polymers of HBV DNA have shown clearly that the 2.1-kb RNA is the messenger for the major protein of the envelope and the 3.5-kb RNA the messenger for the core protein^{42,43}. The mRNAs encoding the other viral proteins remain unknown, but the 3.5-kb RNA probably encodes the DNA polymerase.

Other transcripts have been characterized by *in vitro* experiments: two polymerase II-dependent L(-)-strand transcripts, initiated respectively at positions 1,680 and 2,810, and one polymerase III-dependent S(+) transcript, initiated at position 1,630 and 700 bases long^{44,45}. The two first transcripts could be used for translation of the pre-S1 and pre-C regions. Virus-specific poly(A)⁺ transcripts were also characterized in hepadna animal viruses (Fig. 3). For WHV and GSHV, two major transcripts of 2.1 and 3.7 kb plus additional minor transcripts were found in infected liver with active viral replication^{46,110}. For DHBV, a third major poly(A)⁺ transcript, initiated a few nucleotides downstream from the initiation codon of the pre-S1 region, was detected⁴⁷. Note that for both WHV and DHBV, the 2.1-kb RNA initiates downstream from the initiation codon of the pre-S2 region. This observation raises the question of the nature of the transcript responsible for the translation of the pre-S2 region.

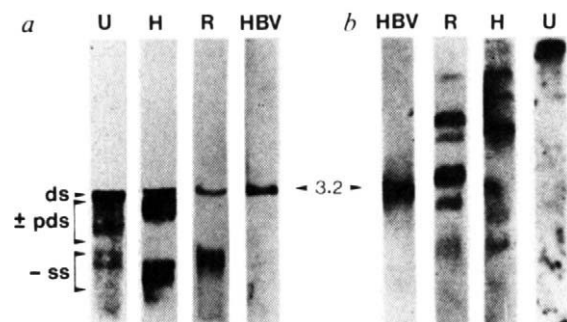


Fig. 4 State of HBV DNA in the hepatocyte. HBV-infected liver DNA was analysed by Southern blotting using HBV DNA as a probe. *a*, Free replicative forms detected in the liver from a chronic carrier. ds, Double-stranded DNA; $\pm$ pds, partially double-stranded DNA; -ss, minus-strand single-stranded DNA. *b*, integrated HBV sequences detected in a liver tumour from a patient with HCC. U; undigested DNA; H, *Hind*III digest; R, *Eco*RI digest; HBV, cloned HBV DNA.

Concerning the regulatory elements which control HBV gene transcription, Cattaneo *et al.*⁴⁰ have noted the presence of a sequence resembling the late promoter of simian virus 40 (SV40) in the pre-S1 region at position 3,122. This sequence may direct the synthesis of the 2.1-kb RNA. No particular sequence which could be the promoter of the 3.5-kb RNA has been noted. On the contrary, a 'TATA'-like promoter sequence situated at an appropriate distance upstream of the 3.7-kb RNA start site was found in WHV and DHBV^{46,110}. The polyadenylation signal common to both mRNAs is located at position 1,916, 20 bp upstream from the 3' end^{40,41}. Little information is available concerning the control mechanisms of viral transcription. It is clear that the mechanisms controlling the synthesis of the 2.1- and 3.5-kb RNAs are different. This is based on the observation that, in the liver during viral replication, the 2.1- and 3.4-kb RNAs are present in about the same amount. On the other hand, in patients with integrated viral sequences, mostly HBsAg is produced, and in mouse cells containing integrated HBV DNA, the amount of 2.1-kb RNA is at least 10 times higher than that of the 3.5-kb RNA⁴². This suggests that the S gene promoter directs constitutive expression of the S gene whereas transcription from the C gene promoter seems to be inducible. Recently, a transcriptional enhancer element was detected 450 bp upstream of the C gene promoter⁴⁸. This element exhibits a preferred activity in human hepatocytes. The tissue specificity of this regulatory element could in part explain the low level of expression of the 3.5-kb RNA observed in the transfected mouse cells. In addition, in transgenic mice containing integrated HBV DNA, S gene transcription takes place preferentially in the liver (C. Pourcel, unpublished data). A tissue-specific enhancer element may also be involved in the regulation of S gene transcription.

Multiplication cycle of HBV

HBV DNA in the hepatocyte can exist in two states, free or integrated into the host cellular chromosome (Fig. 4)^{49,50}. Free HBV DNA, which represents intermediate forms of replication, is detected during the acute and some chronic stages of HBV infection. Integrated sequences are mostly observed during chronic virus infection and especially in HCC. Free and integrated forms can coexist within the same sample but this does not mean that both states coexist in the same cell.

HBV DNA has also been detected in non-hepatic tissues such as kidney, pancreas and skin^{51,52}. Extra-hepatic free viral sequences were also found in the pancreas and kidney of infected

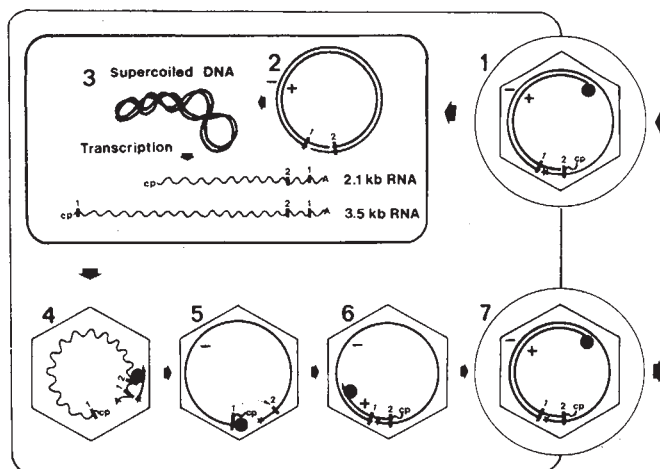


Fig. 5 Replication cycle. (1) Entry into the cell of an extracellular virion containing the HBV genome with a short capped RNA (cp), the DNA polymerase (●) and the DNA-linked protein (*). DR1 and DR2 are indicated by 1 and 2. (2) Complete double-stranded open circular DNA. (3) Supercoiled covalently closed circular DNA. (4) Pregenome packaged into a nucleocapsid. (5) core containing the complete (-) strand DNA. (6) Synthesis of (+) strand. (7) Coated nucleocapsid giving rise to extracellular virion.

Pekin ducks⁵³. The presence of free and integrated forms of HBV DNA in bone marrow and mononuclear blood cells during both the acute and chronic stages of HBV infection raises questions about the role of these cells in the pathogenesis of hepatitis B^{54,55}. It has been suggested that infection of mononuclear cells would modify the immune response⁵⁶.

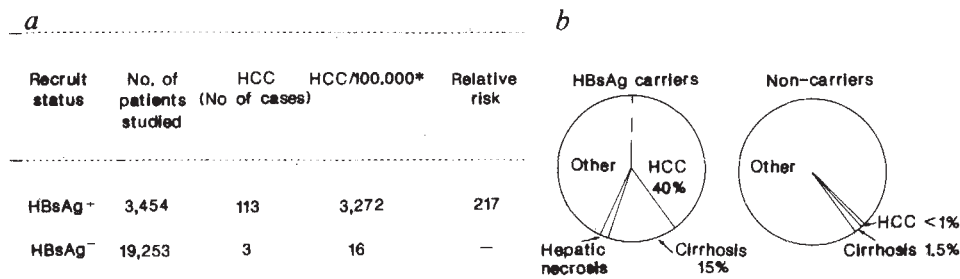
The replication mechanism of hepatitis B viruses, discovered by Summers and Mason for DHBV^{57,58} and confirmed later for HBV⁵⁹⁻⁶², differs strikingly from that of other DNA viruses. The central feature is the use of a RNA copy of the genome as intermediate in replication. The replication cycle involves a reverse transcription step resembling that of retroviruses.

In Pekin ducks 6 hours after infection, DHBV DNA is found in the liver, in both supercoiled and relaxed circular forms. Slightly later, RNA is detected, followed by the appearance of single-stranded DNA⁵⁸.

The HBV replication cycle is believed to consist of the following steps. After penetration of the virus into the hepatocyte, the DNA reaches the nucleus and is converted to complete open circular double-stranded DNA and then to supercoiled DNA (Fig. 5). The minus strand is transcribed by the cellular RNA polymerase to the 3.5-kb RNA previously described and called the pre-genome. Multiple copies of the pre-genome are synthesized from each genome. The pre-genome is encapsidated with the DNA polymerase and the DNA-linked protein. The minus-strand DNA is then synthesized by reverse transcription of the pre-genome. Initiation of reverse transcription takes place at DR1 and the DNA-linked protein covalently bound to the 5' end of the minus strand probably acts as a primer for minus-strand synthesis⁶³. During DNA synthesis, the pre-genome is degraded by an RNase H-like activity and the product of reverse transcription is a full-length minus DNA strand. A small RNA fragment of about 20 bp, probably derived from the 5' end of the pre-genome, is then used to prime the synthesis of the plus strand at DR2 (J. M. Lien and W. Mason, personal communication). Complete synthesis of this strand is not necessary for coating of the nucleocapsids and export of virus particles. This explains why infectious particles contain DNA with a single-stranded region.

The replication cycle of hepadna viridae in some ways resembles that of retroviruses and of CaMV. All these viruses involve a reverse transcription step, but whereas the virion of retroviruses contains RNA and the intermediate form of replication is integrated DNA, the virion of hepadna viruses and CaMV contains DNA and the intermediate replication form is RNA.

Fig. 6 Epidemiology of hepatocellular carcinoma¹. *a*, Relative risk of HCC in relation to recruitment HBsAg status. *b*, Cause of death in relation to recruitment HBsAg status.



* During 6.2 years follow-up.

Integrated sequences have never been detected during DHBV replication (C. E. Rolger and J. Summers, personal communication). An integration step of the genome is therefore probably not necessary for hepadna virus replication. Minus-strand synthesis of retrovirus and CaMV DNA is primed by a transfer RNA while hepadna DNA minus-strand synthesis is probably primed by a protein. The order of the three retroviral genes, *gag*, *pol*, *env*, can be compared to that of the order of the hepadna virus genes C, P and S, but whereas retrovirus proteins are translated from mRNA produced by splicing from a single primary transcript, hepadna proteins are translated from at least two unspliced primary transcripts. For all three classes of virus, promoters are closely followed by polyadenylation signals that are passed over during the first transcription run.

HBV and hepatocellular carcinoma

Evidence that HBV causes HCC comes from epidemiological studies. There is a striking worldwide geographical correlation between the prevalence of HBsAg chronic carriers and the annual incidence of HCC⁶⁴. Moreover, retrospective epidemiological studies have shown that HBsAg is more frequently observed in the serum of patients with HCC than in the control population⁶⁴. The strongest evidence for the role of HBV in HCC comes from prospective epidemiological studies performed in Taiwan by Beasley *et al.*⁶⁵, who studied male subjects aged between 40 and 59 y. A male population was chosen because HCC is about three times more frequent in males than females. The results showed that the relative risk to HBsAg carriers of developing HCC was 217 compared with non-carriers (Fig. 6). Furthermore, 51% of deaths among HBsAg carriers were caused by cirrhosis or HCC compared with only 2% among the control population. Among the non-carriers, 90% possessed an HBV marker, anti-HBs and/or anti-HBc antibodies. The high incidence of HCC is therefore clearly related to the HBsAg chronic carrier state and not to HBV infection *per se*.

We can conclude that although the data do not preclude a role for other factors (such as aflatoxin) in the development of HCC, additional factors other than HBV chronic infection are not necessary to explain the epidemiological studies. It is also of interest that HCC developed generally after a history of cirrhosis over a long period (30–40 y) of chronic carriage. However, additional evidence linking HBV and HCC comes from the WHV/woodchuck model. It has been possible to induce HCC experimentally in woodchuck by inoculation with WHV at birth (J. Gerin, personal communication). This result provides experimental support for an oncogenic role of WHV and validates the use of WHV as a pathogenic model for HBV.

The structure of the HBV DNA sequences present in HBsAg-positive HCC has been studied by Southern blotting. Both human HCC-derived cell lines and human tumour samples^{49,50,66–72} have been analysed. In both cases, HBV DNA sequences integrated into the host genome were nearly always detected. For any given tumour sample, the number of integration sites varied from 1 to 12. The presence of only free viral DNA without integrated sequence, as observed in the chronic carrier stage, has not been reported for HCC. However, free HBV DNA with intermediate forms of replication in addition to the integrated sequences are occasionally observed in the early stage of tumour growth. Virus replication, as well as virus

gene expression, are generally absent in the poorly differentiated hepatocytes of advanced tumours⁷³. It is not known why HBV replication and expression should be absent from HCC cells. Two hypotheses can be envisaged; either cells harbouring HBV yet without replication/expression are selected during tumour growth or the replication/expression decreases gradually in proportion to the dedifferentiation of the cell. HBcAg has been suggested as the target for the cytotoxic immune response⁷⁴. If so, hepatocytes unable to express HBcAg would escape the immune response and be selected during carcinogenesis.

To investigate further the role of HBV in HCC, the presence of HBV DNA in HBsAg-negative HCC was determined. In a significant number of cases, integrated HBV DNA sequences were detected even in patients without any HBV serological marker^{75–78}. These data offer new evidence for an oncogenic role of HBV and indicate that integration may play a part in tumour formation; however, the exact significance of integrated HBV sequences is unclear. Indeed, HBV sequences are not indicative of tumorous tissue. HBV DNA was observed both in non-tumorous parts of the livers with HCC and in chronic hepatitis or cirrhosis without HCC^{50,79}. In addition, when the tumorous and the non-tumorous parts of liver were analysed simultaneously, the integration patterns were generally different but in a very few cases the tumorous and non-tumorous tissues shared the same pattern.

The structure of integrated HBV DNA sequences present in tumour samples^{80,81} and in HCC-derived cell lines^{82–85} was determined by DNA cloning. Viral sequences can be complete genomes or subgenomic fragments that are often rearranged. Such complex rearrangements have already been described for integrated viral sequences of SV40 and adenovirus^{86,87}. The mechanism of HBV integration remains unclear; however, Dejean *et al.*⁸¹ have reported the existence of a virus-specific site of integration. In two independent liver tumours, one of the two host-virus DNA junctions mapped within either DR1 or DR2. The HBV DNA structure which precedes the integration event is unknown. If the substrate of integration is linear or an open circle, integration through the DR sequence could suggest integration at the ends of the HBV DNA strands. If the substrate of integration is a covalently closed circular form, the result observed implies a specific recombination event within the DR. C. Shih (personal communication) reported similar results and found in two hepatoma samples, virus-cell and virus-virus junctions within a few nucleotides of a DR sequence. Taken together, these data show that HBV DNA can integrate via a specific viral DNA sequence. It is nevertheless possible that a nonspecific mechanism of integration could occur in addition to that involving DR1 and DR2. No evidence for a specific cellular integration site or for extensive sequence homology between viral and cellular DNA at this site has been reported to date.

The role of HBV in HCC is not understood at the molecular level. HBV may act as an initiator by inducing continuous liver necrosis and regeneration. Integration could occur at random and cells capable of escaping the immune response would become targets for further events of carcinogenesis. By contrast, HBV integration may have a direct role in liver cancer. However, none of the classical mechanisms of viral oncogenesis has so far been demonstrated. The existence of a viral transforming gene would imply the existence of a conserved viral sequence

in the different tumorous cells. No peculiar region of the HBV genome was conserved in the different clones of integrated HBV sequences in human HCCs⁸¹⁻⁸⁴ and integrated WHV sequences in woodchuck HCCs⁸⁸. In addition, transformation experiments with HBV DNA have been unsuccessful. It is also possible that a hybrid or partial gene product could induce transformation. Recently, Freytag von Loringhoven *et al.*⁸⁹ characterized hybrid HBV-host transcripts in the hepatoma-derived cell line PLC/PRF5. The slow transforming retroviruses, which do not contain a viral oncogene, act as insertional mutagens. The integration of the provirus can activate the expression of a neighbouring cellular oncogene or disrupt a cellular regulator gene, preventing its expression. The two models require that the viral integration occur at approximately the same site in the host genome. Thus far, there is evidence neither for HBV integration in a distinct chromosomal region nor for integration adjacent to any of the known cellular oncogenes. Further characterization of integrated HBV sequences and their flanking sequences in HCC might allow us to find a common element between the different integration sites and hence clarify the process of HBV oncogenesis.

Recombinant DNA and a future vaccine

The worldwide distribution of HBV infection with the consequence of chronic liver disease and HCC implies the need for an efficient and safe vaccine. Compared with more conventional antiviral vaccines, the hepatitis B vaccine is unique in that HBV cannot be propagated in tissue culture. The vaccine is obtained from plasma of healthy HBV chronic carriers. It consists of purified defective 22-nm HBsAg particles formulated in an alum adjuvant. Vaccination strategies differ according to the level of endemicity of the affected population. In countries with low endemicity such as in Western Europe or North America, vaccination is restricted to individuals with an established high risk of acquiring hepatitis B. These populations are mainly health personnel, dialysis patients, drug addicts, male homosexuals and infants of HBsAg-positive mothers. On the other hand, in developing countries with high endemicity such as in Far-East Asia and tropical Africa, a programme of universal vaccination at infancy and early childhood is desirable, if not currently feasible.

The effectiveness and safety of the present hepatitis B vaccine have been demonstrated both in high-risk adult populations⁹⁰⁻⁹² and in newborn infants⁹³⁻⁹⁵. Nevertheless, alternative methods for vaccine production are needed. This is because of limitations in the availability of human serum, the need for elaborate purification and inactivation procedures and the requirement of an inocuity test in chimpanzees. Plasma-derived vaccines are very expensive and therefore beyond the financial possibilities of many countries which need mass infant vaccination programmes.

Skelly *et al.*⁹⁶ have prepared polypeptides in water-soluble micelle form from HBsAg particles. This subunit vaccine induces antibodies at a higher titre than HBsAg particles but still relies on the availability of human sera. For this reason, the use of synthetic oligopeptides carrying HBsAg determinants is an attractive alternative. Unfortunately, the low value of the immune response so far obtained¹³⁻¹⁸ limits the use of this technique. Within this context, development of recombinant DNA vaccines is of great importance. Several approaches can be used and the choice of the technique depends on the knowledge of the structure and the immunogenicity of the antigenic determinants present on the surface of the envelope, mainly HBsAg but also pre-S-encoded epitopes. HBsAg is a conformational antigen and full immunogenicity of HBsAg is highly dependent on its tertiary structure¹¹. Unlike HBcAg⁹⁷, HBsAg is not assembled as particles in bacteria and only bacterium-HBV hybrid polypeptides showing HBsAg antigenicity were obtained⁹⁸⁻¹⁰⁰. A eukaryotic system such as mammalian cell¹⁰¹ or yeast¹⁰² capable of correctly synthesizing native HBsAg particles is therefore required.

At present, the most advanced recombinant vaccine is a vac-

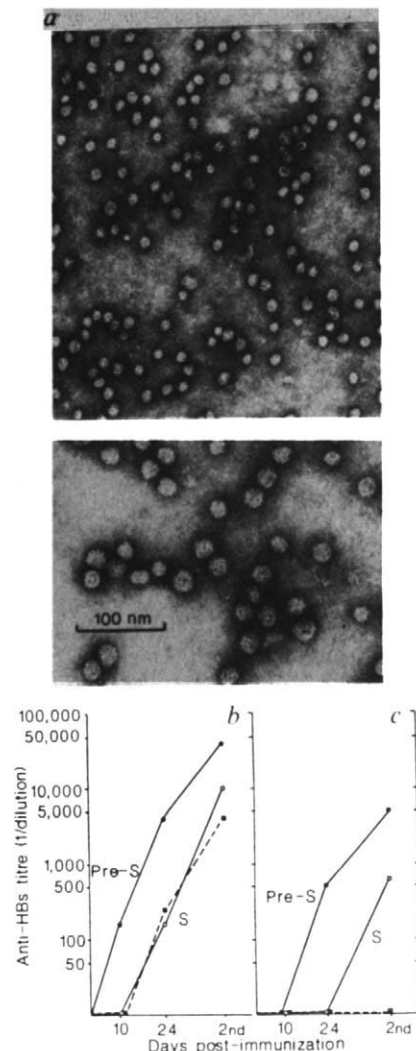


Fig. 7 Immunogenicity of HBsAg 22-nm particles containing the pre-S2 region translation product¹⁰⁷. *a*, Electron micrograph of HBsAg 22-nm particles synthesized in CHO cells. *b*, Production of antibodies to HBsAg (S) and the pre-S2 region translation products (pre-S) in a HBsAg responder mouse strain (strain B10). *c*, Production of antibodies in a HBsAg non-responder mouse strain (strain B10S). Dashed curves, anti-HBs response of mice immunized with HBsAg particles lacking the pre-S2 region determinants. Mouse immunizations were at days 0 and 24. The term '2nd' on the x axis indicates the secondary immune response.

cine produced in yeast. Valenzuela *et al.*¹⁰², using an expression vector containing the S gene placed under the control of the yeast alcohol dehydrogenase I promoter, obtained the synthesis of HBsAg particles in *Saccharomyces cerevisiae*. Miyanohara *et al.*¹⁰³ obtained a comparable result using the acid phosphatase promoter. The particles had the same density as particles from human serum but were heterogeneous in size; they contained the major protein only in its non-glycosylated form. Injected into chimpanzees, these HBsAg particles induced anti-HBs antibodies that conferred immunity to HBV infection¹⁰⁴. Human trials are now in progress.

Production of HBsAg 22-nm particles in mammalian cells is a second alternative; the particles obtained are glycosylated and secreted in the cell supernatant¹⁰¹. An increase in HBsAg production was obtained by the following method^{105,106}. A plasmid containing the cDNA encoding dihydrofolate reductase (*dhfr*) and HBV sequences was introduced into Chinese hamster ovary (CHO) cells lacking *dhfr*. Selection of *dhfr*-positive clones resistant to methotrexate resulted in gene amplification of both *dhfr* cDNA and HBV sequences, and an increase of HBV DNA transcription. Using this procedure, Michel *et al.*¹⁰⁵ have constructed a plasmid containing the S gene and the pre-S2 region

placed under the control of both the SV40 early promoter and the promoter localized in the pre-S1 region. The HBsAg 22-nm particles obtained (Fig. 7) contained both the major protein in its two forms, glycosylated and non-glycosylated, and the glycosylated middle protein in a ratio comparable to that of virion. The particles, homogeneous in size, contained both the HBsAg determinants and the pHSA receptor activity, as a result of the presence of the pre-S2 region translation product. Studying the murine immune response to these particles Milich *et al.* showed that the pre-S2 region product is significantly more immunogenic than HBsAg and that immunization of a HBsAg non-responder strain with these particles can circumvent non-responsiveness to HBsAg (Fig. 7)¹⁰⁷. Because the appearance of anti-pHSA receptor antibodies seems to be a highly reliable criterion for viral clearance, the HBsAg particles obtained in CHO cells may constitute a particularly efficient vaccine.

The third approach for a recombinant vaccine is to construct a live vaccine, using as vector a virus capable of replicating in humans. Smith *et al.* have constructed a live vaccine using a vaccinia virus as vector¹⁰⁸. The S gene was placed under the control of the vaccinia virus early promoter. Cells infected with the vaccinia virus recombinants synthesized and excreted

HBsAg particles; infected rabbits produced antibodies to HBsAg. Chimpanzees infected with the recombinant vaccine did not develop an anti-HBs response but were protected against hepatitis B following challenge with HBV¹⁰⁹.

Which of these three hepatitis B recombinant vaccines will constitute for the near future the most appropriate alternative to the present hepatitis B vaccine? This question is not easily answered because the final choice of vaccine depends on several parameters: the immunogenicity of the particles, the possibilities of an industrial transfer of the technical procedure and the safety of the final product. In any case, twenty years after the fundamental discovery of HBsAg by B. S. Blumberg¹¹¹, the future recombinant DNA vaccine against hepatitis B will be the first example of the application of genetic engineering to human vaccination as well as the first example of a vaccine against an infectious disease related to cancer.

The nomenclature used for HBV proteins and genes is that agreed upon a Cold Spring Harbor meeting (May 1985).

We thank Kallee Gehring, Simon Wain-Hobson and Christian Br  chet for revising this manuscript. This work was supported by NIH grant CA97300-02, the Economic European Community and the Association pour la Recherche contre le Cancer.

1. Beasley, R. P. & Hwang, L. Y. in *Viral Hepatitis and Liver Disease* (ed. Vyas, G. N.) 209-224 (Grune & Stratton, New York, 1984).
2. Chisari, F. V. (ed.) *Advances in Hepatitis Research* (Masson, New York, 1984).
3. Robinson, W. S., Marion, P., Feitelson, M. & Siddiqui, A. in *Viral Hepatitis 1981 int. Symp.* (eds Szmuness, W., Alter, H. J. & Maynard, J. E.) 57-68 (Franklin Institute Press, Philadelphia, 1982).
4. Summers, J., Smolek, J. & Snyder, R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4533-4537 (1978).
5. Marion, P. L., Oshiro, L. S., Regnery, D. C., Scullard, G. H. & Robinson, W. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2941-2945 (1980).
6. Mason, W. S., Seal, G. & Summers, J. *J. Virol.* **35**, 829-836 (1980).
7. Courou  c-Pauty, A. M., Plancon, A. & Soulier, J. P. *Vox Sang.* **44**, 197-211 (1983).
8. Stibbe, W. & Gerlich, W. H. *J. Virol.* **46**, 626-628 (1983).
9. Heermann, K. H. *et al.* *J. Virol.* **52**, 396-402 (1984).
10. Tiollais, P., Wain-Hobson, S. in *Advances in Hepatitis Research* (ed. Chisari, F. V.) 9-20 (Masson, New York, 1984).
11. Mishiro, S. *et al.* *J. Immun.* **124**, 1589-1593 (1980).
12. Peterson, D. L., Nath, N. & Gavilanes, F. J. *biof. Chem.* **257**, 10414-10420 (1982).
13. Lerner, R. A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **78**, 3403-3407 (1981).
14. Dreesman, G. R. *et al.* *Nature* **295**, 158-160 (1982).
15. Prince, A. M., Ikram, H. & Hopp, T. P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 579-582 (1982).
16. Bhatnagar, P. K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **79**, 4400-4404 (1982).
17. Neurath, A. R., Kent, S. B. H. & Strick, N. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7871-7875 (1982).
18. Gerin, J. L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **80**, 2365-2369 (1983).
19. Neurath, A. R., Kent, S. B. H. & Strick, N. *Science* **224**, 392-394 (1984).
20. Neurath, A. R., Kent, S. B. H., Strick, N., Taylor, P. & Stevens, C. E. *Nature* **315**, 154-156 (1985).
21. Machida, A. *et al.* *Gastroenterology* **86**, 910-918 (1984).
22. Albin, C. & Robinson, W. S. *J. Virol.* **34**, 297-302 (1980).
23. Petit, M. A. & Pillot, J. *J. Virol.* **53**, 543-551 (1985).
24. Tiollais, P., Charnay, P., Vyas, G. N. *Science* **213**, 406-411 (1981).
25. Ohori, H., Shimizu, N., Yamada, E., Onodera, S. & Ishida, N. *J. gen. Virol.* **65**, 405-414 (1984).
26. Takahashi, K. *et al.* *J. Immun.* **130**, 2903-2907 (1983).
27. Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P. & Charnay, P. *Nature* **281**, 646-650 (1979).
28. Pasek, M. *et al.* *Nature* **282**, 575-579 (1979).
29. Valenzuela, P., Quiroga, M., Zaldivar, J., Gray, P. & Rutter, W. J. in *Animal Viral Genetics* (eds Fields, B., Jaenisch, R. & Fox, C. F.) 57-70 (Academic, New York, 1980).
30. Ono, Y. *et al.* *Nucleic Acids Res.* **11**, 1747-1757 (1983).
31. Fujiyama, A. *et al.* *Nucleic Acids Res.* **11**, 4601-4610 (1983).
32. Kobayashi, M. & Koike, K. *Gene* **30**, 227-232 (1984).
33. Tiollais, P. *et al.* in *Viral Hepatitis and Liver Disease* (ed. Vyas, G. N.) 49-65 (Grune & Stratton, New York, 1984).
34. Toh, H., Hyashida, H. & Miyata, T. *Nature* **305**, 827-829 (1983).
35. Moriarty, A. M., Alexander, H. & Lerner, R. A. *Science* **227**, 429-433 (1985).
36. Kay, A., Mandart, E., Treppe, C. & Galibert, F. *EMBO J.* **4**, 1287-1292 (1985).
37. Galibert, F., Chen, T. N. & Mandart, E. *J. Virol.* **41**, 51-65 (1982).
38. Seeger, C., Ganem, D. & Varmus, H. E. *J. Virol.* **51**, 367-375 (1984).
39. Mandart, E., Kay, A. & Galibert, F. *J. Virol.* **49**, 782-792 (1984).
40. Cattaneo, R., Will, H., Hernandez, N. & Schaller, H. *Nature* **305**, 336-338 (1983).
41. Cattaneo, R., Will, H. & Schaller, H. *EMBO J.* **3**, 2191-2196 (1984).
42. Pourcel, C. *et al.* *J. Virol.* **42**, 100-105 (1982).
43. Gough, N. J. *molec. Biol.* **165**, 683-699 (1983).
44. Rall, L. B., Standing, D. N., Laub, O. & Rutter, W. J. *Molec. cell. Biol.* **3**, 1766-1773 (1983).
45. Standing, D. N., Rall, L. B., Laub, O. & Rutter, W. J. *Molec. cell. Biol.* **3**, 1774-1782 (1983).
46. M  ry, T., Etienne, J., Tr  pe, C., Tiollais, P. & Buendia, M. A. *EMBO J.* **4**, 1507-1514 (1985).
47. B  scher, M., Reiser, W., Will, H. & Schaller, H. *Cell* **40**, 717-724 (1985).
48. Shaul, Y., Rutter, W. J. & Laub, O. *EMBO J.* **4**, 427-430 (1985).
49. Br  chet, C. *et al.* *Hepatology* **2**, 26S-34S (1982).
50. Shafritz, D. *Hepatology* **2**, 35S-41S (1982).
51. Dejean, A., Lugassy, C., Zafrani, S., Tiollais, P. & Br  chet, C. *J. gen. Virol.* **65**, 651-655 (1984).
52. Siddiqui, A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4861-4864 (1983).
53. Halpern, M. S. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **80**, 4865-4869 (1983).
54. Rom  t-Lemonne, J. L. *et al.* *Science* **221**, 667-669 (1983).
55. Pontisso, P., Poon, M. C., Tiollais, P. & Br  chet, C. *B. med. J.* **288**, 1563-1566 (1984).
56. Elfassi, E. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **81**, 3526-3528 (1984).
57. Summers, J. & Mason, W. S. *Cell* **29**, 403-415 (1982).
58. Mason, W. S., Aldrich, C., Summers, J. & Taylor, J. M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3997-4001 (1982).
59. Blum, H. E., Haase, A. T., Harris, J. D., Walker, D. & Vyas, G. N. *Virology* **139**, 87-96 (1984).
60. Fowler, M. J. F., Monjardino, J., Tsiuay, K. N., Zuckerman, A. J. & Thomas, H. C. *J. med. Virol.* **13**, 83-91 (1984).
61. Miller, R. H., Tran, C. T. & Robinson, W. S. *Virology* **139**, 53-63 (1984).
62. Miller, R. H., Marion, P. L. & Robinson, W. S. *Virology* **139**, 64-72 (1984).
63. Molnar-Kimber, K. L., Summers, J. W. & Mason, W. S. *J. Virol.* **51**, 181-191 (1984).
64. Szmuness, W. *Prog. med. Virol.* **24**, 40-69 (1978).
65. Beasley, R. P., Hwang, L. Y., Lin, C. C. & Chien, C. S. *Lancet* vol. 2 1129-1133 (1981).
66. Marion, P. L., Salazar, F. H., Alexander, J. J. & Robinson, W. S. *J. Virol.* **33**, 795-806 (1980).
67. Chakraborty, P. R., Ruiz-Opazo, N., Shouval, D. & Shafritz, D. A. *Nature* **286**, 531-533 (1980).
68. Br  chet, C., Pourcel, C., Louise, A., Rain, B. & Tiollais, P. *Nature* **286**, 533-535 (1980).
69. Edman, J. C., Gray, P., Valenzuela, P., Rall, L. B. & Rutter, W. J. *Nature* **286**, 535-538 (1980).
70. Chen, D. S., Hoyer, B. H., Nelson, J., Purcell, R. H. & Gerin, J. L. *Hepatology* **2**, 42S-45S (1982).
71. Shafritz, D. A., Shouval, D., Sherman, H. I., Hadziyannis, S. J. & Kew, M. C. *New Engl. J. Med.* **305**, 1067-1073 (1981).
72. Hino, O. *et al.* *Hepatology* **4**, 90-95 (1984).
73. Hirohashi, S., Shimoto, Y., Ino, Y. & Kishi, K. *J. natn. Cancer Inst.* **69**, 565-568 (1982).
74. Edleston, A. L. W. F., Mondelli, M., Mieli-Vergani, G. & Williams, R. *Hepatology* **2**, 112S-127S (1982).
75. Br  chet, C. *et al.* *New Engl. J. Med.* **306**, 1384-1387 (1982).
76. Figs, A. *et al.* *Hepatology* **4**, 364-368 (1984).
77. Hino, O., Kitagawa, T. & Sugano, H. *Int. J. Cancer* **35**, 5-10 (1985).
78. Br  chet, C. *et al.* *New Engl. J. Med.* **312**, 270-276 (1985).
79. Br  chet, C. *et al.* *Lancet* **ii**, 765-768 (1981).
80. Dejean, A., Brechet, C., Tiollais, P. & Wain-Hobson, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2505-2509 (1983).
81. Dejean, A., Sonigo, P., Wain-Hobson, S. & Tiollais, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5350-5354 (1984).
82. Koshy, R. *et al.* *Cell* **34**, 215-223 (1983).
83. Shaul, Y. *et al.* *J. Virol.* **51**, 776-787 (1984).
84. Mizusawa, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 208-212 (1985).
85. Ziemer, M., Garcia, P., Shaul, Y. & Rutter, W. J. *J. Virol.* **53**, 885-892 (1985).
86. Botchan, M., Topp, W. & Sambrook, J. *Cell* **9**, 269-287 (1976).
87. Sambrook, J. *et al.* *Cold Spring Harb. Symp. quant. Biol.* **44**, 569-584 (1980).
88. Ogston, C. W., Jonak, G. J., Rogier, C. E., Astrin, S. M. & Summers, J. *Cell* **29**, 385-394 (1982).
89. Freytag von Loringhoven, A., Koch, S., Hofschneider, P. H. & Koshy, R. *EMBO J.* **4**, 249-255 (1985).
90. Crosnier, J. *et al.* *Lancet* **i**, 455-459 (1981).
91. Laplanche, A. *et al.* *Lancet* **i**, 222 (1982).
92. Szmuness, W. *et al.* *New Engl. J. Med.* **307**, 1481-1486 (1982).
93. Maupas, P. *et al.* *Lancet* **i**, 289-292 (1981).
94. Lee, G. C. Y. *et al.* *J. infect. Dis.* **148**, 526-529 (1983).
95. Prozesky, O. W. *et al.* *J. infect. Dis.* **7**, Suppl. 1, 53-57 (1983).
96. Skelly, J., Howard, C. R. & Zuckerman, A. J. *Nature* **290**, 51-54 (1981).
97. Cohen, B. J. & Richmond, J. E. *Nature* **296**, 677-678 (1982).
98. Charnay, P., Gervais, M., Louise, A., Galibert, F. & Tiollais, P. *Nature* **286**, 893-895 (1980).
99. Edman, J. C., Hallowell, R. A., Valenzuela, P., Goodman, H. M. & Rutter, W. J. *Nature* **291**, 503-506 (1981).
100. Fujisawa, Y. *et al.* *Nucleic Acids Res.* **11**, 3581-3590 (1983).
101. Dubois, M. F. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **77**, 4549-4553 (1980).
102. Valenzuela, P., Medina, A., Rutter, W. J., Ammerer, G. & Hall, B. D. *Nature* **298**, 347-350 (1982).
103. Miyahara, A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1-5 (1983).
104. McAleer, W. J. *et al.* *Nature* **307**, 178-180 (1984).
105. Michel, M. L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **81**, 7708-7712 (1984).
106. Patzer, E. J. *et al.* in *Viral Hepatitis and Liver Disease* (ed. Vyas, G. N.) 477-485 (Grune & Stratton, New York, 1984).
107. Milich, D. R. *et al.* *Science* **228**, 1195-1199 (1985).
108. Smith, G. L., Mackett, M. & Moss, B. *Nature* **302**, 490-495 (1983).
109. Moss, B., Smith, G. L., Gerin, J. L. & Purcell, R. H. *Nature* **311**, 67-69 (1984).
110. Enders, G. H., Ganem, D. & Varmus, H. *Cell* (in the press).
111. Blumberg, B. S., Alter, H. J. & Visnitch, S. J. *Am. med. Ass.* **191**, 541-546 (1965).

Tectonics of the Qinling Belt: build-up and evolution of eastern Asia

M. Mattauer*, Ph. Matte*, J. Malavieille*, P. Tapponnier†, H. Maluski*,
Xu Zhi Qin‡, Lu Yi Lun§ & Tang Yao Qin§

* Department des Sciences de la Terre, Équipe Tectonique, Université des Sciences et Techniques du Languedoc, Place E. Bataillon, 34060 Montpellier Cédex, France

† Institut de Physique du Globe, 4, Place Jussieu, 75230 Paris Cédex 05, France

‡ Institute of Geology, Chinese Academy of Geological Sciences, Baiwanzhuang Road, Beijing, China

§ Geological Bureau of Shaanxi Province, Xian, China

The structure and history of the Qinling range, which separates the North and South China cratons, are a key to the tectonic evolution of eastern Asia. The mountain range is composed of a northern Palaeozoic belt and a southern Mesozoic belt. The Palaeozoic orogeny probably involved the closure of an oceanic basin along a north-dipping subduction zone, followed by south-vergent obduction of ophiolites and intracontinental thrusting during the Devonian. The Mesozoic belt appears to be the product of purely intracontinental, shallow crustal deformation over a large decollement with at least 100 km of horizontal transport. Since the Devonian, both belts have been sliced by huge east-west sinistral strike-slip faults.

THE east to west striking Qinling mountain belt (eastern China) separates the Sino-Korean Craton (North China Block, NCB) and the Yangtse Craton (South China Block, SCB) for >1,000 km (refs 1–5; Fig. 1). To the west, this belt connects with the Qilian and Kun Lun belts. To the east, it appears to be truncated by the north-south striking Tan Lu Fault. Knowledge of the structure and history of the Qinling belt is important to place constraints on the way in which the eastern part of the Asian continent was assembled and evolved during the Phanerozoic. Considerable work by Chinese scientists⁶ suggests that the Qinling mountains are a polyphase orogen juxtaposing a narrow Palaeozoic belt to the north with a broader Lower Mesozoic belt to the south^{7–15}. Both belts are intruded by mostly Carboniferous to Triassic granites and are sliced by large E-W vertical faults. Some mafic and ultramafic rocks^{16,17} have been interpreted as the ophiolitic remnants of a Palaeozoic suture. Although the Qinling belt has been frequently studied in detail, there is still no agreement on various evolutionary models in the framework of plate tectonics. In contrast to the Palaeozoic closure proposed above^{15,16}, for instance, a recent synthesis based on stratigraphic correlations¹⁸ implies a Lower Mesozoic age for the first suturing between NCB and SCB. Similarly, there are problems with the timing and magnitudes of Palaeozoic-Mesozoic and Tertiary movements inferred from palaeomagnetic measurements^{19,20}.

The main purpose of our study was to analyse small-scale deformations along two ~200 km long traverses south-east of Xian (Fig. 2), with particular attention to the kinematics, amounts and age of shortening and shearing. This field work, performed in autumn of 1983 and spring of 1984, was used to establish a tectonic model of the Qinling belt that integrates all regional observations made so far (maps at 1/250,000 and 1/500,000)⁶ as well as some complementary mapping done by us on the basis of Landsat images and 1/65,000 aerial photographs. The constraints provided by this tectonic model on the large-scale movements of NCB relative to SCB since the Palaeozoic are also discussed.

Tectonic zonation of the belt

Within the Qinling belt, several tectonic zones can be defined from the stratigraphy and the age and characteristics of deformation and metamorphism. The zones are generally separated by large late Palaeozoic-Mesozoic to Cenozoic strike-slip faults (Figs 2, 3)²¹.

Zone 1. Low-grade deformation at the southern edge of the Sino-Korean Craton. An old basement consisting mainly of

gneisses and amphibolites (2,300–2,600 Myr) is unconformably overlain by a volcano-sedimentary Proterozoic sequence^{10–12}. Both are unconformably overlain by a thick (>2,000 m) pile of mainly platform sediments (shales, sandstones, tillitoids, limestones, dolomites and conglomerates) from Upper Proterozoic (Sinian) to Cambrian and Ordovician. The deformation is simple in the north with gentle open folds and steep local cleavage. Southwards, folds become more tight and overturned to the south with a strong slaty cleavage and a north-south down-dip stretching lineation, particularly well marked in conglomerates and oolitic limestones. At the southern rim of zone 1, deformation becomes more complex with increasing metamorphism (biotite). It is polyphased in the Ordovician where the early S₁ cleavage is folded by crenulations and large-scale folds with associated S₂ and S₃ cleavages. These three phases of deformation and the metamorphism occurred before the deposition of undeformed coal-bearing Permian strata. Mesozoic granitic batholiths cross-cut all these terranes.

Zone 2. Intense synmetamorphic thrusting and recumbent folding. This unit shows a much stronger Palaeozoic deformation related to higher grade metamorphism (biotite to kyanite)^{11,12}. The Proterozoic volcano-sedimentary and sedimentary series belonging to NCB have been transformed into a thick slab (several kilometres) of micaschists, garnet-bearing gneisses, amphibolites and siliceous marbles. On a small scale, deformation is polyphased with an S_{1–2} northward-dipping foliation and

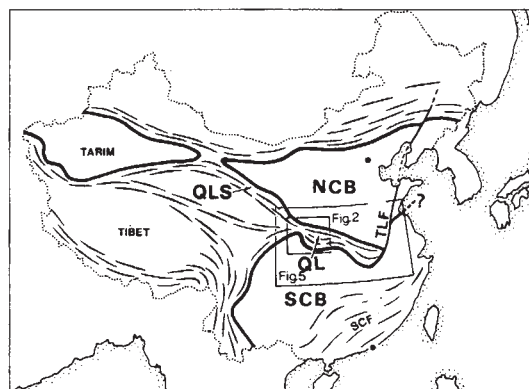


Fig. 1 General setting of the Qinling belt between NCB and SCB. QL, Qinling belt; QLS, Qilian Shan; SCF, South China Caledonian fold belt; TLF, Tan Lu Fault.

Fig. 2 Simplified structural sketch of the studied area (from the Shaanxi Province 1/500,000 geological map; ref. 6). Circled numbers indicate the different structural zones. Broken lines show the itinerary. Box: Palaeozoic belt: (1) Archaean basement of the Sino Korean Craton; (2) Lower and Upper Sinian; (3) Palaeozoic sediments and ophiolites (ω). CO, Cambro-Ordovician; D, Devonian; (4) Palaeozoic granites. Mesozoic belt: (5) Proterozoic basement of the Yangtse Craton; (6) Sinian volcanics; (7) Upper Sinian to Mesozoic sediments; (8) Mesozoic granites; (9) slaty cleavage; (10) foliation. (11) Mesozoic folds; (12) thrusts; (13) strike-slip faults; (14) Tertiary; and (15) location of geochronological samples. L.f., Lo Nan fault; SX.f., Shang Xian fault; DF.f., Deng Fung fault; SY.f., San Yan fault.

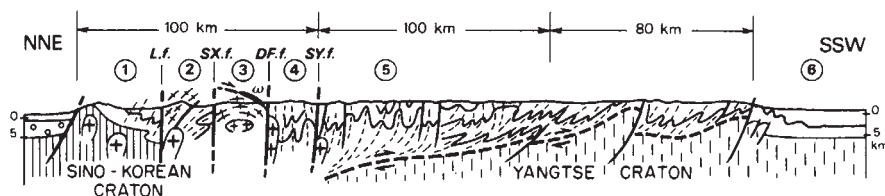
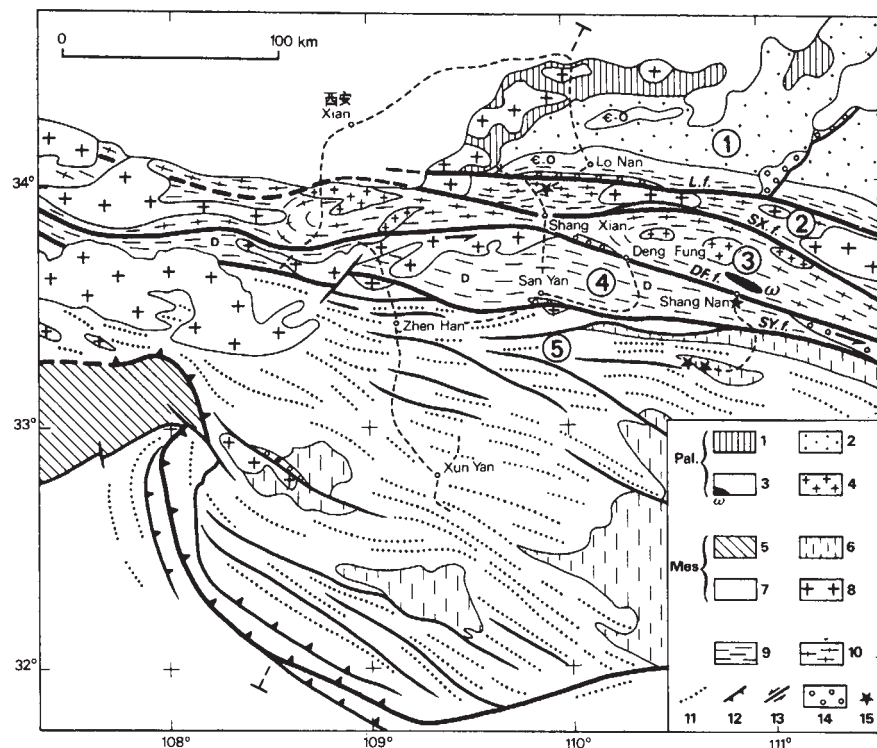


Fig. 3 General simplified cross-section of the Qinling belt. Vertical scale exaggerated. Circled numbers refer to the structural zones described in Fig. 2. Cleavage is indicated by dotted lines.

syn- S_1 to late- S_2 crystallization of muscovite and large patches of biotite. From north to south, that is, from the top to the bottom of the metamorphic pile, the metamorphism decreases in intensity (kyanite to biotite and chlorite) with isograds apparently parallel to the foliation. Such an inverted metamorphism could be related to ductile south-vergent overthrusting, as in the Himalayas²². In the lowermost, least metamorphic part of this zone (biotite, chlorite, muscovite) the horizontal schistosity (S_{1-2}), which bears a N-S stretching lineation, corresponds to the axial plane of large recumbent folds. The presence of upside-down stromatolites in the Sinian marbles implies the existence of flat inverted limbs up to 5 km wide from north to south. As in the north, all the metamorphic rocks are intruded by Mesozoic granites.

Zone 3. Central Qinling domes. This zone corresponds to the most metamorphic, presumably deeper, central part of the Palaeozoic belt of the Qinling¹¹. A previously gneissic horizontal foliation corresponding perhaps also here to recumbent folds and thrusts is folded in a broad domelike structure. Various generations (syn to post-kinematic) of granitic rocks (white and pink aplites and pegmatites, granites and diorites) were emplaced as sills and dykes parallel to the foliation or cross-cutting it. A large (20 km $\times$ 2 km) massif of peridotite outcrops in the southern part of the dome in which two main structural units can be distinguished. The lower unit consists mainly of micaschists, marbles, mafic tuffs, cherts and amphibolites, some of which show chemical oceanic affinities¹⁷. Palaeozoic foraminifera have been found laterally in corresponding, less metamorphic terranes⁶. The upper unit is a thick slab of mylonitic chromite-rich harzburgite (2–3 km), overlain by garnet-bearing amphibolites and an alternation of micaschists, fine-grain amphibolites and amphibolite gneisses which may represent metabasalts and metatuffs. This unit, of clear oceanic nature, is strongly deformed with a N-S stretching lineation

underlined in the amphibolites by coronitic shadows around the garnets. Such coronitic reactions could correspond to those observed in some retromorphic eclogites. Following Chinese authors, we interpret this mafic-ultramafic unit as an incomplete, allochthonous ophiolitic slab. As the thrusts are strongly refolded, it is difficult to determine the vergence of thrusting.

Nevertheless, a study of preferred orientations of minerals in thin sections, and electron microscope observations of dislocations in olivines of the mylonitic harzburgites indicate that both hot and cold strains are compatible with south-directed shear (Xu, Zh. Q. in preparation). The nature of the autochthonous basement in the domes is unknown but the great amount of granitic rocks emplaced in them suggests that it probably consists of continental crust. The early granites appear to be of Variscan age¹¹, but the age of the climax of metamorphism and deformation remains unknown. The southern boundary of zone 3 shows evidence of late ductile deformation with prominent E-W horizontal lineations probably in relation to sinistral strike-slip shear.

Zone 4. Devonian trough. This zone is characterized by a thick Devonian sequence (up to 4,000 m) consisting mainly of slates, sandstones, greywackes, polygenic conglomerates and some limestones, locally overlain by continental Carboniferous^{6,14}. The sequence is strongly folded by steep axial planar folds with a fan-like cleavage, which locally refold more ancient recumbent folds. In the northern part a strong horizontal E-W stretching lineation develops along vertical fault zones which can be interpreted as sinistral strike-slip faults. Away from these faults, boudinage and strain markers indicate down-dip stretching. The metamorphism reaches the amphibolite facies with high temperature/low pressure characteristics, producing hornfels in pelitic rocks and scapolites in marbles. In the Middle Devonian between Dan Feng and Shang Nan, we found conglomerates with pebbles of gabbros, peridotites, amphibolites, granitic

rocks, quartz, garnet and amphibole clasts. Such crystalline pebbles could come from zones to the north and correspond to erosion of mafic and ultramafic as well as continental rocks of a deep part of the lower Palaeozoic belt now absent at the longitude of our traverse but perhaps located much more to the west.

Zone 5. Lower Mesozoic 'Indosinian' fold belt. In this area a 10,000-m-thick pile of Upper Sinian to Triassic sediments (mainly limestones, sandstones and slates), is strongly folded⁶. The upper part of the pile (Devonian to Triassic) corresponds to a spectacular belt of E-W-striking flexural-slip folds of multi-kilometric wavelength with steep limbs up to 2 km thick. Steeply dipping ESE faults are also present. An incipient S_1 steep to fanlike cleavage develops in the most incompetent layers (cleavage in slates, dissolution cleavage in marly limestones). The strike and dip of the cleavage are rather constant whereas the plunge of the intersection lineations varies considerably (from horizontal to vertical), suggesting the existence of variable bedding dips before folding. The stretching lineation deduced from strain markers, pressure shadows and boudinage is down-dip. In the lower part of the pile (Sinian to Silurian), the attitude of the folds and cleavage is different. S_1 is sub-horizontal or shallow dipping to the north with a strong N-S stretching lineation. Similar isoclinal F_1 folds are overturned or recumbent towards the south.

Bedding parallel shear zones with south-vergent displacements are observed within the Sinian limestones together with steep axial planar folds F_2 (homoaxial with F_1), and a vertical crenulation cleavage. The latter may be coeval with the large-scale domes in the lower Sinian volcanics. The transition from steep fan-like cleavage in the upper part of the folded sedimentary pile to horizontal foliations in the lower part is gradual. As in the Rhenohercynian belt of Europe²³, the passage from cleavage to foliation is typically concave upwards. The horizontal attitude of the foliation and stretching lineation at depth may be related to ductile décollements within the Lower Palaeozoic and Sinian series. The main décollement surface probably lies under the competent Sinian limestones in the upper part of the deeper and more ductile acid volcanics. Progressive shortening has apparently resulted in superimposed folding ($F_1 + F_2$) in the deeper, horizontally foliated part and in fanning of S_1 and local northwards overturning of the folds in the upper part of the sedimentary pile where S_1 was initially vertical.

The metamorphism, which increases downwards, is marked by the appearance of chloritoid, phengite, garnet and biotite in slates. At the boundary between the Upper Sinian limestones and the underlying acidic and basic volcanics of the Lower Sinian (ignimbrites, keratophyres, mafic tuffs and basalts), various occurrences of blue amphiboles previously identified as glaucophane¹³, appear to be riebeckite in the samples we analysed with the microprobe.

Geochronology

Although many radiometric ages on rocks of the Qinling belt have now been published in the Chinese literature¹¹, difficulties in evaluating the analytical procedures by which such ages were obtained made it necessary to perform new measurements. The first quantitative results of this work are summarized here.

The $^{39}\text{Ar}/^{40}\text{Ar}$ technique has been applied to pure metamorphic minerals. Selection of the minerals was conducted to answer questions about the age of the metamorphism, in zones 2, 4 and 5. Biotite and muscovite extracted from the micaschists of zone 2 both give well-defined plateau ages at 328 ± 7 and 348 ± 7 Myr, respectively (Fig. 4). Such ages reflect 'Variscan' metamorphism in this area. The difference in age of 20 Myr between the biotite and the muscovite is roughly within the uncertainty on each age measurement. It could also be related to the better retention of argon in muscovites than in biotites, the latter remaining an open system for argon while muscovite still accumulated the radiogenic product.

South-west of Shang Nan, in zone 4, Devonian micaschists with subvertical foliation and E-W stretching lineation have

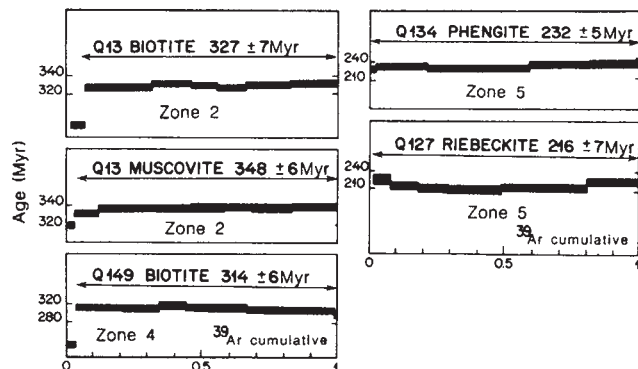


Fig. 4 Age spectra for analysed minerals. Location of samples shown in Fig. 2.

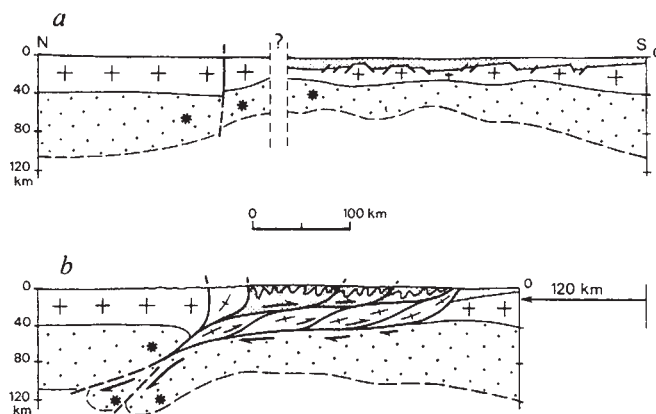


Fig. 5 Interpretative lithospheric scale section illustrating the mechanism of continental subduction. *a*, Possible situation before subduction showing the thinned Yangtze Plate; *b*, aspects of folding and décollements related to continental subduction.

suffered also an 'Hercynian' metamorphism. The age plateau, defined by six increasing temperature steps (representing >95% of the ^{39}Ar released) indicates an age of 314 ± 6 Myr (Fig. 2).

The other minerals that we have dated so far originate in zone 5 (Indo-Sinian fold belt). Phengites (Si-Al substitution of 2.3) yield a well-defined age plateau of 236 ± 5 Myr and riebeckites nearby seem to be slightly younger (217 ± 8 Myr). Despite the slight discrepancy between the latter two ages, the growth of these minerals appears to be clearly related to the Indosinian orogeny.

Left lateral movements

Many E-W vertical faults⁹⁻¹² cut the Qinling belt; the most important ones correspond to boundaries between the above-mentioned structural zones. All appear to be strike-slip faults with probably large cumulated offsets. Most were active during the Tertiary and some are still active²⁴. Along them, a brittle phase of deformation often overprints older ductile shear.

From the outline of the regional geology and tectonics given here, the Den Fung Fault appears to be the most important. We could follow it for over 300 km, but it is seen on the 1/500,000 geological maps and Landsat images to continue eastwards and westwards, reaching a total length of ~1,000 km. A strong mylonitic vertical foliation develops along it in schists, gneisses, granitoids and diorites with characteristic lens-shaped sigmoidal 'fishes', horizontal slickensides and stretching lineations, indicative of left lateral movement. The corresponding shear zone may reach 1,000 m in thickness in some areas (for example, along the road from Xian to Zhen Han and east of Den Fung). Total offset on this fault could exceed a few hundred kilometres. The age obtained on biotites in the Devonian micaschists of zone 4 implies that part of the ductile deformation could be Carboniferous in age.

Three other major fault zones also appear to correspond to

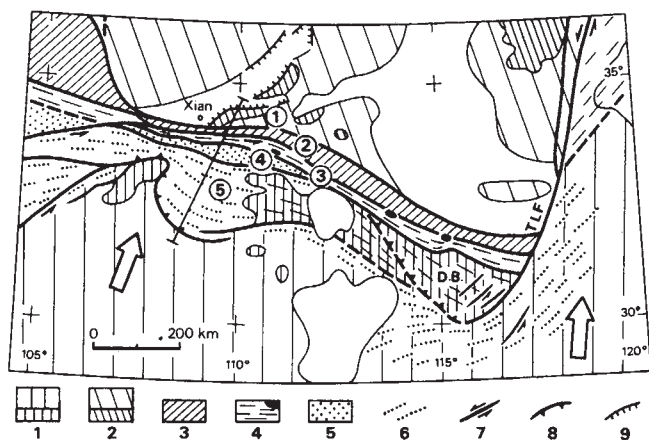


Fig. 6 Structural scheme of the whole Qinling belt showing the location of the studied section. Circled numbers refer to structural zones: (1) Yangtse Craton with basement+Sinian volcanics (Lower) and Upper Sinian to Mesozoic cover (Upper); (2) Sino Korean Craton with basement (Lower) and cover (Upper); (3) zone 2; (4) zone 3 with ophiolites; (5) Devonian trough. (6) Mesozoic folds in zone 5; (7) strike-slip faults; (8) thrusts; (9) normal faults. Arrows indicate the displacement of the Yangtse relative to the Sino-Korean Craton.

large left-lateral strike-slip zones: the Lo Nan fault, which separates zones 1 and 2; the Shang Xian fault, between zones 2 and 3, along which vertical shreds of Carboniferous, Triassic and Jurassic are smeared; and the San Yan fault, between zones 4 and 5.

All these strike-slip faults must have strongly modified the original geometry of the Palaeozoic and Lower Mesozoic belt and testify for large left-lateral movements between the Yangtse and the Sino-Korean cratons since the Palaeozoic.

Evolution of the Palaeozoic belt

Despite the Mesozoic and Tertiary tectonic imprints, it is possible to reconstruct tentatively the existence and closure of a Palaeozoic ocean and Caledonian orogeny. The presence of a large slice of ultramafic to mafic rocks (peridotites and amphibolites) overlain by a thick Palaeozoic volcano-sedimentary sequence suggests the existence of an oceanic basement in a basin located close to a continent. The microfossils found in the cherts⁶ interbedded within the volcano-sedimentary series indicate a lower Palaeozoic age for this basin (Figs 5, 6).

The oceanic crust was subsequently thrust over marbles and amphibolites overlying a probable continental crust. The N-S stretching lineations can be interpreted as a result of shearing during south-vergent obduction. The hot deformation in the harzburgites implies that the obduction was coeval with the demise of a north-dipping subduction zone. The absence of calc-alkaline volcanism in this part of the Qinling mountains suggests that movements along the strike-slip faults may have separated the corresponding magmatic arc from the deformed zone observed here.

The age of the main synmetamorphic deformation (post-obduction and pre-dome) in zone 3 is still unknown. In this more internal part of the belt, it could correspond to a Caledonian event. The presence of various crystalline (ultramafic, mafic and sialic) pebbles in the middle Devonian of zone 4, which can be considered as a molassic foredeep of the Palaeozoic belt, supports this inference.

The Variscan tectonics have apparently involved three major events. Intracontinental shortening produced folding in zone 1, and large-scale southwards synmetamorphic thrusting in zone 2, probably between 350 and 320 Myr BP. In zone 3, an episode of widespread granitization followed the above tangential tectonics. Chinese geochronological studies¹¹ imply Variscan ages for some of these granites. The end of the Hercynian evolution of the belt may have involved important movements along the ductile strike-slip shear zones which separate zone 3 from zone

4. Some support for this inference is found in the age (~314 Myr) of deformations in vertically foliated and horizontally stretched Devonian rocks not far from the fault zone.

Evolution of the Mesozoic fold belt

The evolution of the Mesozoic (Indosinian) belt appears to have been simpler. The thick Upper Sinian to Triassic folded sequence is the sedimentary cover of the Yangtse Craton. No clear non-conformity can be seen between the Lower Sinian volcanics and the Upper Sinian limestones. The significance of this volcanism is still debated. The thick overlying sedimentary series corresponds to various depositional conditions. The Upper Sinian and Lower Palaeozoic (5–6 km) were apparently deposited on a passive margin bordering a Palaeozoic ocean to the north. The corresponding platform was cut by many faults. Extension is marked in the southernmost part of the studied area by alkaline to per-alkaline Ordovician magmatism. It is thus probable that the continental crust of the Yangtse Craton was considerably thinned since the Upper Proterozoic. The Devonian and Carboniferous platform type deposits are also very thick (up to 4 km), but locally show abrupt variations in thickness, perhaps in relation to synsedimentary normal faults recorded by local conglomerates and limestone breccias along the edges of horsts and grabens (Figs 5, 6). In zone 4 the continental character of the Carboniferous may be related to the Variscan orogeny to the north. By this time the Yangtse and Sino Korean cratons had joined together and strike-slip movements along the Deng Fung Fault may have started. The Triassic is represented by thick platform limestones overlain by molasse. The subsequent Mesozoic folding thus affected a thinned but clearly continental crust (Figs 3, 5, 6).

The downwards transition from steep to recumbent folding implies the existence of a major 'décollement' at depth. The observation of horizontal foliations in Lower Sinian domes over a large area, from north to south, shows the large extent of the décollement surface. Unfolding the upper competent layers on detailed 1/50,000 sections, suggests a minimum shortening of ~40% in zone 5, over a present width of ~100 km and ~35% shortening farther south over a width of ~80 km. The total displacement thus obtained on the décollement is ~100 km. As usual, this value may be considered a lower bound. The folded zone may be compared to an 'accretionary wedge' made of the sediments detached from the Yangtse basement in a process akin to subduction (Figs 3, 5; see refs 23, 25 for continental subduction). As the same style of folding is observed to the east (Dabie Shan), a similar décollement may occur there. Farther to the east the belt is truncated by the large N-S left-lateral Tan Lu strike-slip fault²⁷. Displacement on the Tan Lu Fault may account for the arcuate trend of Indosinian folds (E-W to NE-SW), perhaps also in relation to décollements at depth (Fig. 6).

Relative movements between SCB and NCB

It is worthwhile to compare the constraints on relative movements between SCB and NCB provided by the structural evolution outlined here with the results of recent palaeomagnetic studies^{19,20}. Palaeomagnetic results imply that NCB and SCB were widely separated in the Cambrian, because their palaeopoles are quite different²⁰. This is in agreement with the presence of ophiolites of Lower Palaeozoic age in the Qinling belt. Devonian conglomerates in zone 4 contain pebbles that came from a probable Caledonian belt then situated not far north. The lack of oceanic suture in zones 4 and 5 concurs to indicate that SCB was close to NCB at that time, as proposed by Zhang *et al.*¹.

Independently of palaeomagnetic data^{19,20}, our geological data suggest that SCB used to be located much farther west relative to NCB than it is at present. One important piece of geological evidence is that the source of the crystalline pebbles in the Devonian (zone 4) may have been the Caledonian belt, which is well developed 800 km farther west in the Qilian Shan. Another is that NCB and SCB are now separated by large

sinistral ductile strike-slip faults.

We therefore believe that since the first collision between NCB and SCB in the Devonian, the two cratons have remained in contact, with only intracontinental deformation including, in particular, large strike-slip movements. This view, which converges with that of Zhang *et al.*¹, is at variance with those of McElhinny¹⁹ and Klimetz¹⁸ who have proposed that NCB and SCB were separated by an ocean during the Permian. Our study suggests that movements between NCB and SCB since the Devonian have only involved major sinistral strike-slip movements, probably in the Carboniferous, followed by convergence roughly normal to the Qinling belt with intracontinental subduction and shortening during Upper Triassic times. The geometry of the Triassic belt and particularly the presence of the N-S, sinistral Tan Lu Fault at its eastern end is hard to reconcile with important strike-slip movements at this time. The final Tertiary and Quaternary movements between NCB and SCB also occurred along E-W sinistral strike-slip faults.

Although the discrepancy between Ordovician to Triassic poles of Lin *et al.*²⁰ does imply eastward relative movements

between SCB and Siberia, and thus possibly NCB, these authors suggest that this motion occurred mostly during the Tertiary²⁰. The timing they propose is therefore quite different from that outlined above. However, more recent assessments of palaeomagnetic evidence both in SCB and NCB, as well as in Korea and Japan^{28,29}, indicate that such motions cannot have occurred after the Jurassic, and thus are more in keeping with our geological conclusions.

This work was carried out within the Research Project of the Chinese Ministry of Geology, Beijing Institute of Geology, Chinese Academy of Geological Sciences and French Centre National de la Recherche Scientifique and Institut National d'Astronomie et de Géophysique, in cooperation with the Geological Bureau of Shanxi Province. We are particularly indebted to Li Tingdong, Xiao Xucheng, Li Chengda, Zhen Yen Li and G. Aubert. We thank A. Leyreloup, A. Nicolas and F. Boudier for petrological discussions; G. Peltzer, Wang Fang Guo and Zhang Zhitao for field collaborations; and the civil authorities of the Shanxi Province and the people of San Yuan for their help during the catastrophic centenary flood of October 1983.

Received 1 April; accepted 16 August 1985.

1. Zhang, Zh.M., Liou, J. G. & Coleman, R. G. *Bull. geol. Soc. Am.* **95**, 295-312 (1984).
2. Zhang, W. Y. *et al. Tectonic Map of China and its Environs* (Science Press, Beijing, 1983).
3. Li, C. Y. *et al. Tectonic Map of China* (Cartographic, Beijing, 1975).
4. Chinese Academy of Geological Sciences. *Geological Map of China 1/4,000,000* (Cartographic, Beijing, 1973).
5. Chinese Academy of Geological Sciences. *Geological Map of Asia 1/5,000,000* (Cartographic, Beijing, 1975).
6. Geological Maps 1/500,000 (1969); 1/250,000 (Geological Regional Survey and Geological Bureau of Shaanxi Province).
7. Li, C. Y., Zhu, Y. W., Feng, B. C. & Whu, H. C. in *Scient. Pap. on Geol. for Int. Exchange* 174-189 (Publishing House of Geology, Beijing, 1978).
8. Li, C. Y. *Bull. chin. Acad. geol. Sci. Ser. 1* **2**, 11-22 (1980).
9. Wang, H. S., Xu, C. G. & Zhou, Z. G. *Acta Acad. Sinica* **56**, 270-280 (1982).
10. Chang, Q. S. *et al. J. Changchun Geol. Inst.* **4**, 1-28 (1979).
11. Chang, Q. S. *Metamorphic Geology of East Qinling*. (People's Publishing House, Jiling, 1980).
12. Cheng, Y. Q. & Zhang, S. G. *Regional Geol. China* **2**, 1-14 (1982).

13. Wu, H. *Bull. Xi'an Inst. Geol. Min. Res.* **7**, 64-81 (1984).
14. Wang, H. Z. & Qiao, X. F. *Geol. Mag.* **121**(6), 599-614 (1984).
15. Wu, H. Q. *Acta geol. sin.* **54**(3), 195-207 (1980).
16. Yu, G., Tian, G. & Lian, J. *Henan Geol.* **56**-63 (1984).
17. Lin, Q., Fu, G. & Zhang, J. *Henan Geol.* **45**-55 (1984).
18. Klimetz, M. P. *Tectonics* Vol. 2 139-166 (1983).
19. McElhinny, M. W., Embleton, J. J., Ma, X. H. & Zhang, Z. K. *Nature* **293**, 212-216 (1981).
20. Lin, J., Fuller, M. & Wen, Y. Z. *Nature* **313**, 444-449 (1985).
21. Breddin, H. *Von. Geol. Mitt.* **12**, 81-130 (1973).
22. Le Fort, P. *Am. J. Sci.* **A275** 1-44 (1975).
23. Bally, A. W. in *Thrust and Nappe Tectonics* 13-32 (Geol. Soc. Lond. 1981).
24. Peltzer, G., Tapponnier, P., Zhang Zhitao & Xu Zhi Qin *Nature* (submitted).
25. Mattauer, M. *Geol. Soc. Lond. Spec. Pap. on Collision Tectonic* (in press).
26. Xu, J. W. in *Scientific Papers on Geology for International Exchange*, 143-162 (Publishing House of Geology, Beijing, 1980).
27. Xu, Zh. Q., Zhang, Q. D. & Zhao, M. *Bull. Chin. Acad. geol. Sci.* 17-44 (1982).
28. Lee, G., Besse, J., Courtillot, V. & Montigny, R. *J. geophys. Res.* (submitted).
29. Courtillot, V. & Besse, J. *Nature* (submitted).

Neogene and Quaternary faulting in and along the Qinling Shan

G. Peltzer*, P. Tapponnier*, Zhang Zhitao†, & Xu Zhi Qin‡

* Institut de Physique du Globe, 4 place Jussieu, 75230 Paris Cédex 05, France

† Institute of Geology and Mineral Resources, Xian, China

‡ Institute of Geology, Chinese Academy of Geological Sciences, Baiwanghuang Road, Beijing, China

The structure of the Qinling Shan mountain range (Shaanxi, China) involves major east-south-east strike-slip faults. Field observations made during two expeditions in 1983 and 1984 show that the faults are active and may have undergone several tens of kilometres of post-Eocene left-lateral displacement. The faults cut the morphology sharply and offset left-laterally small channels. Their trace is marked by cataclastic to mylonitic non-coaxial deformation zones at least 50-100 m wide. The Qinling faults allow the eastward extrusion of south China and may be considered the easternmost continuation of the Altyn Tagh fault system. They link extensional tectonics in northeastern China to the Himalayan collision.

THE Qinling Shan (Shaanxi province, China) is an east-west trending mountain range marking the limit between the North and the South China blocks (Fig. 1). Structurally, the range appears to be formed from the juxtaposition of two parallel orogenic belts¹⁻³: a northern Middle Palaeozoic belt where folding, thrusting and metamorphism resulted from the collision between the Sinokorean and Yangtze platforms, and a southern Lower Mesozoic foldbelt where upright folding of the ≈10,000-m-thick Upper Sinian to Triassic cover of the Yangtze craton occurred above detachments which allowed the continental basement to subduct northwards under the Palaeozoic belt⁴. Large EW-ESE vertical faults separate the main structural zones of the range. Some of them have probably undergone important pre-Tertiary left-lateral displacements^{3,4}.

Despite a history of low seismicity⁵, the study of Landsat images suggests that several of these strike-slip faults are still

active, together with the major normal faults of the Fen Wei graben⁶. We report here first results of field work along several segments of the major faults bounding and cutting the Qinling mountains south-east of Xian, made during a broader structural study of the area³. We also discuss the implications of such results for hypotheses concerning the disruption of the Asian continent by the India/Eurasia collision⁶⁻⁹.

Quaternary displacements and morphology

Inferences drawn from the study of Landsat images (Fig. 2) are confirmed by observations made in the field and by aerial photographs; these provide clear morphological evidence for Quaternary left-lateral displacements on EW-ESE strike-slip faults in the Qinling Shan, as well as for Quaternary throws on normal faults in and along the Wei He graben. A characteristic, straight, narrow furrow runs almost continuously for many

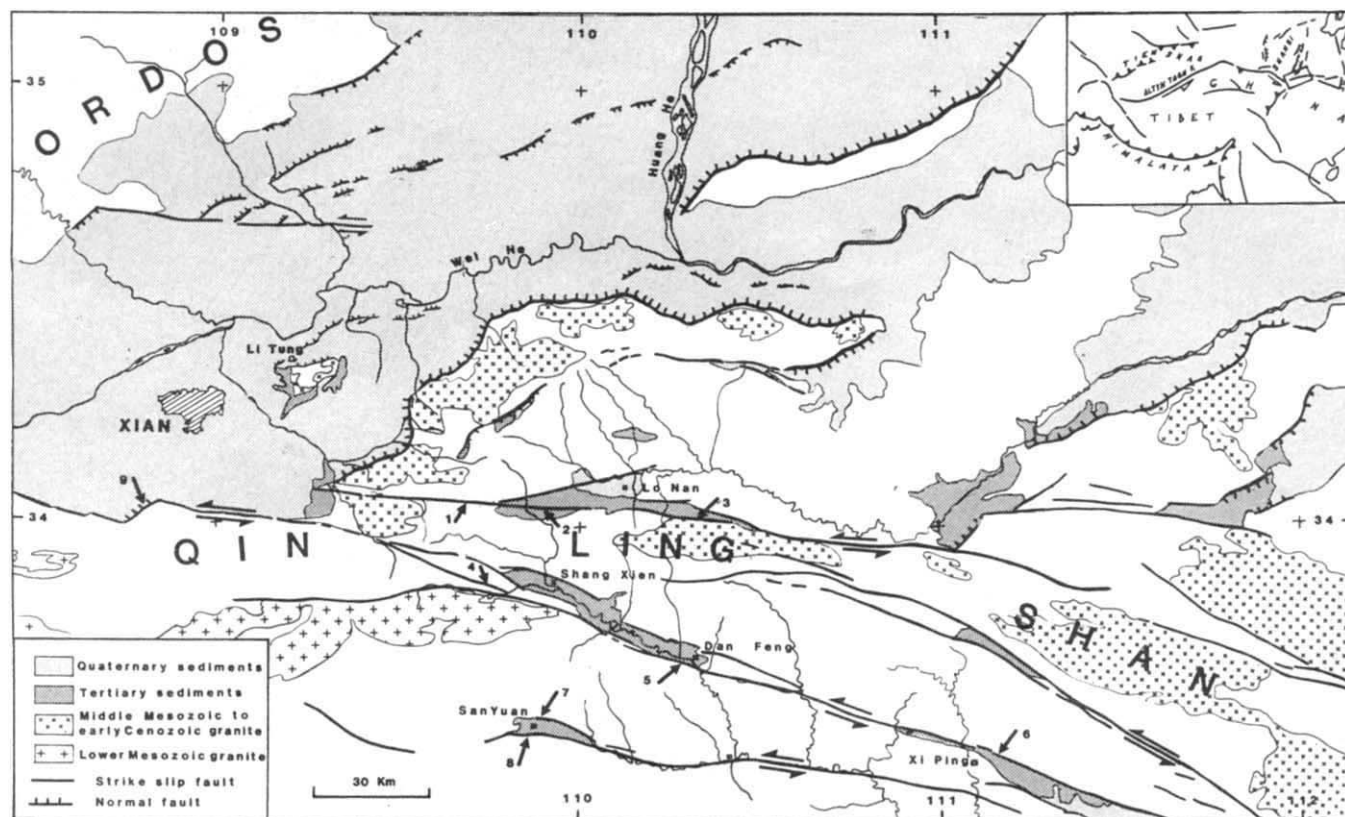


Fig. 1 Simplified map of Cenozoic and Quaternary tectonics in the Qinling mountain belt. Contours are from the geological map of Shaanxi province, 1/500,000. Faults are from Landsat images and aerial photographs complemented by field observations. Numbers and small arrows indicate location of sites and outcrops described in the text.

tens of kilometres along most segments of the Lo Nan and Shang Xien-Dan Feng faults (Figs 1, 2). The sharp rectilinear traces of both faults, particularly remarkable at sites 1, 2, 3, and 5, 6, bound or cut elongated Tertiary and Quaternary basins (Fig. 1). In the field, the furrow seen on the aerial photos generally corresponds to an alignment of small passes and truncated spurs, sometimes outlined by a corridor of cultivated fields. At sites west of Lo Nan, for instance (Figs 1, 3a), it is clear that the fault plane is vertical and that Quaternary displacement on the fault is mostly strike-slip, because no difference in elevation can be observed between the two sides of the fault. Important disruptions of the Quaternary drainage pattern imply late Quaternary left-lateral movements. West of Shang Xien, for instance (Fig. 1, site 4), the trace of the southern branch of the Shang Xien fault is marked by kilometric left-lateral offsets of at least seven adjacent river channels (Fig. 4). A smaller sequence of left-laterally deviated streams is also observed along the Lo Nan fault (Fig. 1, site 1). Some larger rivers, however, do not appear to be offset by the faults, and offsets are not always similar along strike (Fig. 2). Although larger, and thus presumably older, river channels should be more offset than smaller ones, offsets of 10 km or more, equal to or larger than the spacing between major streams draining the Qinling, may be attenuated, cancelled or even reversed by capture¹⁰. Such changes in the river courses will occur more easily in places where the rivers are less deeply entrenched in their channels, such as along the fault-bounded edges of sedimentary basins (for example Xi Ping basin, Figs 1, 2).

Triangular facets several 100 m high indicate important Quaternary normal throws along the San Yan fault (Fig. 1, site 8). Along the large fault which forms the southern edge of the Wei He graben south and east of Xian, Holocene normal throws are also particularly clear. On Fig. 3b, for instance (Fig. 1, site 9), the more gently sloping upper part of the triangular spur is truncated by a steep planar trapezoidal facet, which

suggests a recent increase in the rate of faulting. The relatively small incision made by the stream flowing down the 'wine-glass' valley in the centre of the facet corroborates this inference. North of the Wei He (Figs 1, 2), several north-east striking normal fault scarps, as well as a more linear, probably strike-slip, east-west scarp on which they appear to branch, cut the Quaternary fill of the graben and disrupt the morphology of its floor. These scarps are a good illustration of the combination of strike-slip and normal faulting which characterizes the active and Quaternary tectonics of the whole area.

Deformation near and within fault zones

Rocks were observed to be deformed in a similar brittle way in most places along the large strike-slip faults. In basement rocks as well as in Tertiary terrains, the faults are generally outlined by 50-100-m-wide cataclastic gouge zones in which both the orientation of the steep schistosity and the sense of rotation of phacoid fragments indicate left-lateral strike-slip displacements.

East of Lo Nan, for instance (Fig. 1, site 3), a spectacular 100-m-wide zone of cataclasites and gouges crops out along a river which crosses the fault (Fig. 3c). Within the 110°N striking deformed zone, parallel bands of gouges involving different types of rock are observed: lenses of foliated and stylolitized white marbles with sigmoidal shapes, green and dark-grey clays, reddish-weathered sandstones, dark-brown to black fault gouges, apparently derived from red pelites and shales and fragments of basement rocks such as granites. In all the bands, the steep foliation has a nearly constant strike of 130°N. The non-coaxial deformation suggested by the anticlockwise rotation of numerous phacoids is consistent with left-lateral shear within the gouge zone. This sense of movement is corroborated by the existence of many small left-lateral shears striking ESE-ENE (Fig. 3c), in a manner similar to C and C' planes in ductile shear zones (see ref. 11). The gouge zone limits a Mesozoic granite (Fig. 1) which has suffered brittle to semi-ductile (left-

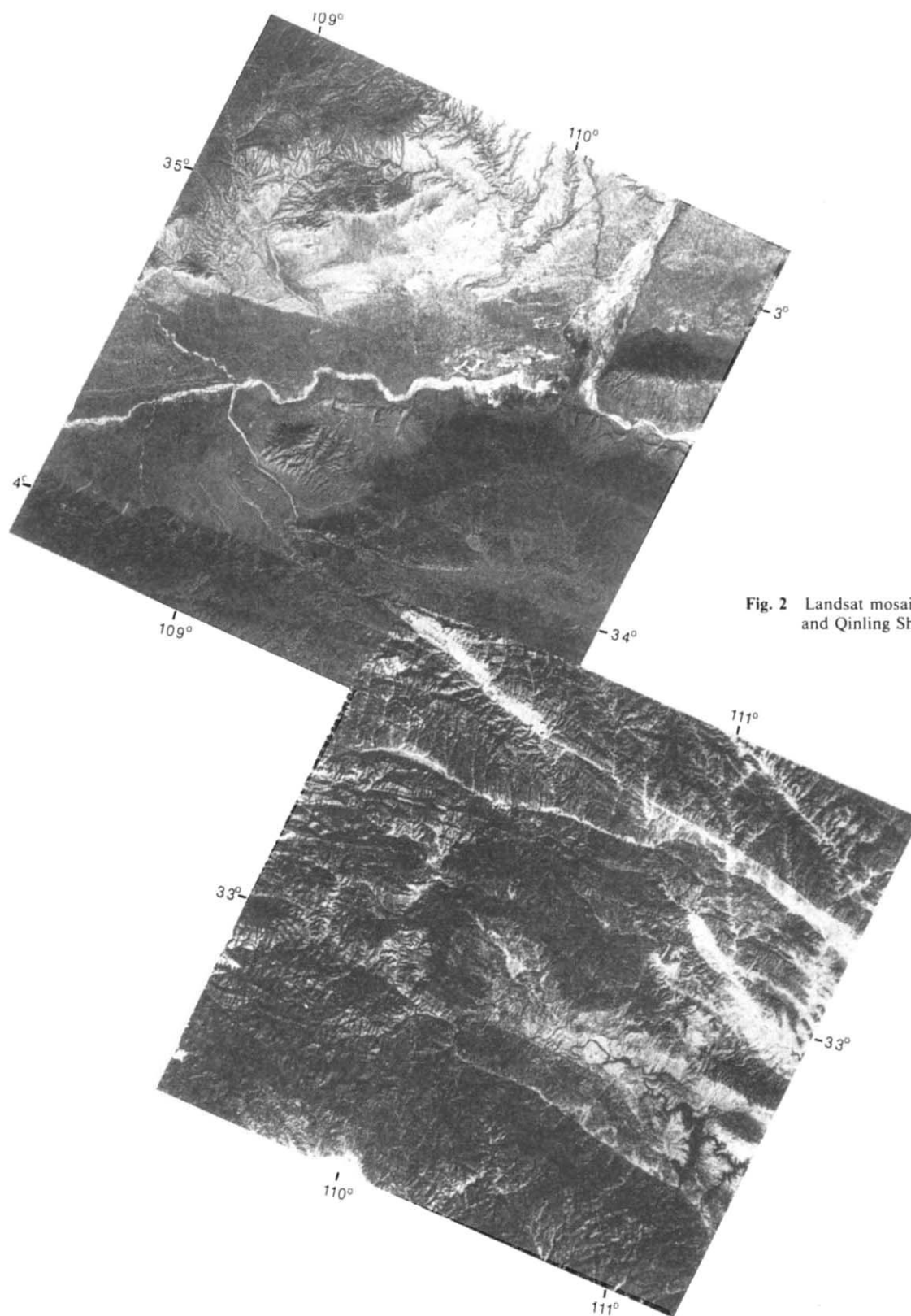


Fig. 2 Landsat mosaic of the Wei He graben and Qinling Shan east of Xian.

lateral) shear in the upper greenschist facies (chlorite) in the vicinity of the fault.

Examples of such brittle deformation are also clear in site 2, where the Lo Nan fault marks the contact between pink Neogene sandstones and pelites to the north and Eocene red conglomerates and sandstones to the south, at the place where the streams bend to the west along the southern branch of the Shang Xien fault (site 4, Fig. 1, and arrowed Fig. 4), and south of Dan Feng (Fig. 1, site 5) where the fault marks the contact between 50°

north-dipping Eocene conglomerate and Devonian slates and greywackes. In this latter site, fractures and cleavage appear in the Eocene rocks several tens of metres away from the fault zone. Closer to the fault, the cleavage becomes an intense cataclastic foliation in which greenish lustrous shear zones, sometimes bearing horizontal slickensides, swing around the largest pebbles. Some of them are crushed into powder. The zone where major displacements have occurred corresponds to a 20-m-wide band of greenish-white gouge where the original

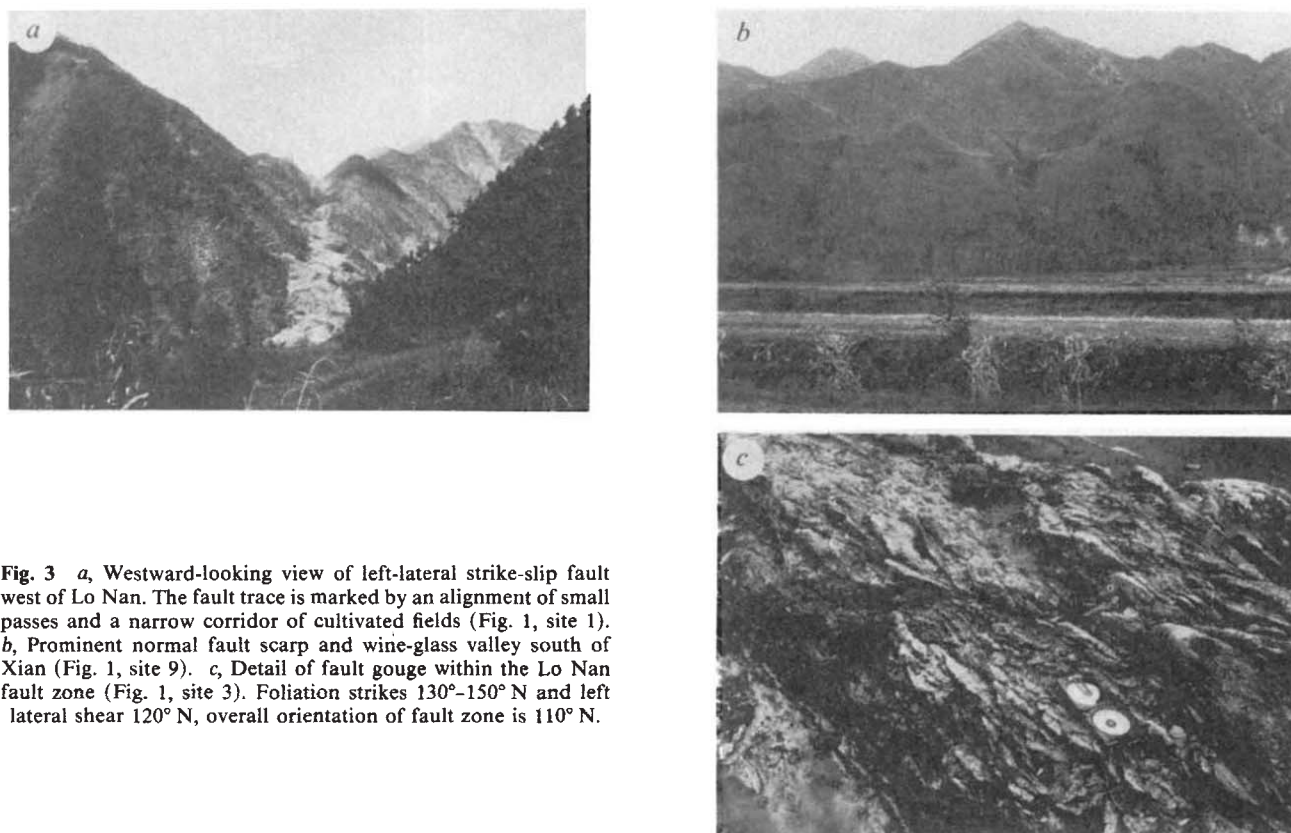


Fig. 3 *a*, Westward-looking view of left-lateral strike-slip fault west of Lo Nan. The fault trace is marked by an alignment of small passes and a narrow corridor of cultivated fields (Fig. 1, site 1). *b*, Prominent normal fault scarp and wine-glass valley south of Xian (Fig. 1, site 9). *c*, Detail of fault gouge within the Lo Nan fault zone (Fig. 1, site 3). Foliation strikes 130° – 150° N and left lateral shear 120° N, overall orientation of fault zone is 110° N.

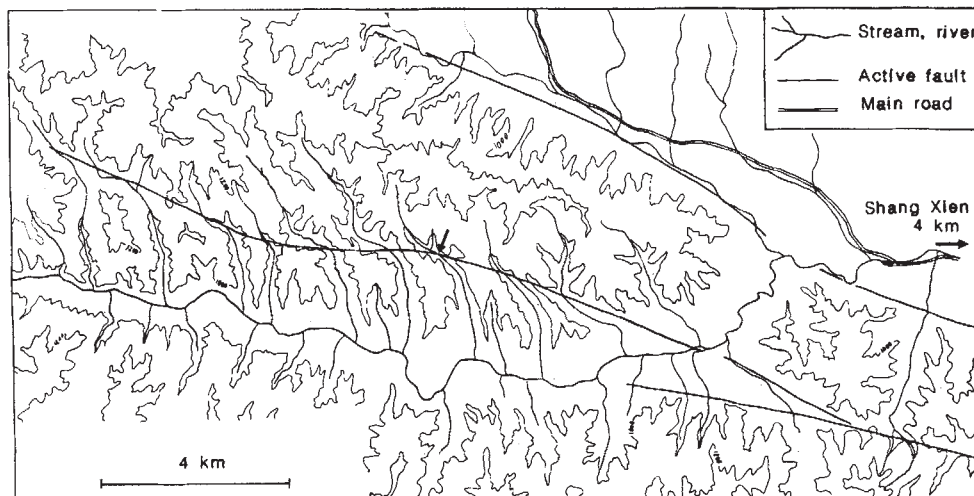
rocks can hardly be recognized.

Within the Eocene-Neogene basins, smaller faults in the red beds (Fig. 1, sites 6, 7) bear slickensides with various pitches and senses, the overall brittle deformation being consistent with left-lateral movements on the major faults. East of Xi Ping, for instance, along the northern edge of the Eocene basin (Fig. 1, site 6), numerous small left-lateral faults striking 110° – 150° N are observed. Most of those striking 140° – 150° N have a significant thrust component. In the San Yan basin, on the other hand (Fig. 1, site 7), slickensides observed on small NE-SE fault planes in the Eocene red beds show a combination of normal and strike-slip movements (Fig. 5a). Although most of the slip vectors shown in Fig. 5b are consistent with a least principal horizontal stress oriented SSE, the northern edge of the basin shows evidence for purely strike-slip right-lateral movements on faults striking 120° – 130° N which are thus incompatible with the slips on Fig. 5b. Clear right-lateral horizontal slickensides are also observed on the main, now largely normal,

San Yan fault plane, east of San Yan. North of San Yan, however, such strike-slip fault planes are always cut by the normal-strike-slip faults shown in Fig. 5b. There is thus a suggestion that, after the deposition of the Eocene red beds, a predominantly strike-slip regime consistent with roughly NNE regional shortening was followed by the more pervasive extensional strike-slip regime illustrated by the measurements in Fig. 5b; the latter regime still dominates the tectonics of the Qinling and Wei He areas, and implies left-lateral components of slip on all faults striking between SSE and ENE, including the major Lo Nan, Shang Xien and San Yan faults.

Along the Lo Nan and Shang Xien faults, the late Cenozoic gouge zones often follow more ductile and locally wider vertical shear zones^{2,3}. Large movements may have occurred within such zones, which generally correspond to boundaries between the main structural units in the Qinling range³. Ductile deformation has produced mylonites with vertical foliations of variable intensity. As in the Tertiary gouge zones, the deformation is

Fig. 4 Left-lateral offsets in the drainage pattern along the Shang Xien fault. The small arrow indicates the site where the gouge was observed. From a topographical map of Shang Xien, 1/100,000 (fig. 1, site 4).



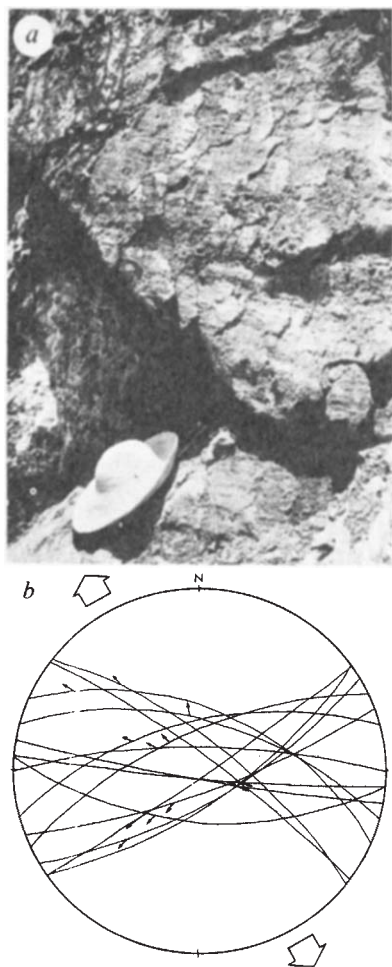


Fig. 5 *a*, North of San Yan, left-lateral slickensides in Eocene sandstones (Fig. 1, site 7). *b*, Slickensides measurements on post-Eocene faults north of San Yuan (lower hemisphere projection) are compatible with roughly SSE extension.

non-coaxial and numerous criteria in the field as well as in thin sections indicate anticlockwise rotation and thus left-lateral shear within the mylonitic zones³. Some of the movements along such ductile zones may have occurred in the Palaeozoic (3), but also during and since the Mesozoic because most of the mesozoic granites are affected by ductile shear along the faults and are transected or offset by them⁴.

Amount of Tertiary displacement

The evidence for late Cenozoic activity on the faults described above is compelling. (1) Tertiary continental deposits are clearly involved in the gouge zones along the faults, and affected by the compatible regional brittle deformation. (2) The strike-slip faults may have controlled the formation of intramontane Tertiary basins, which are elongated adjacent to the faults and could thus represent pull-apart basins along them. (3) The clear morphological traces observed in the field as well as on aerial and satellite photographs, and the offset of the river channels, substantiate late Quaternary activity on the faults.

Although the existence and sense of strike-slip movements in the Qinling Shan since the Eocene can be considered firmly established, the amount of slip experienced by each fault is more debatable. Nevertheless, several points may be emphasized. First, it is unlikely that intensely foliated 50–100-m wide gouge zones containing a great variety of phacoid fragments of basement and cover rocks such as those described above could be the result of displacements smaller than a few tens of kilometres. Second, displacements of 1–1.5 km of the late Quaternary drainage pattern along the Shang Xien fault (Fig. 4) could imply left-lateral slip rates of 1–15 mm yr⁻¹ on that fault alone, if the small offset river channels were 10⁵–10⁶ yr

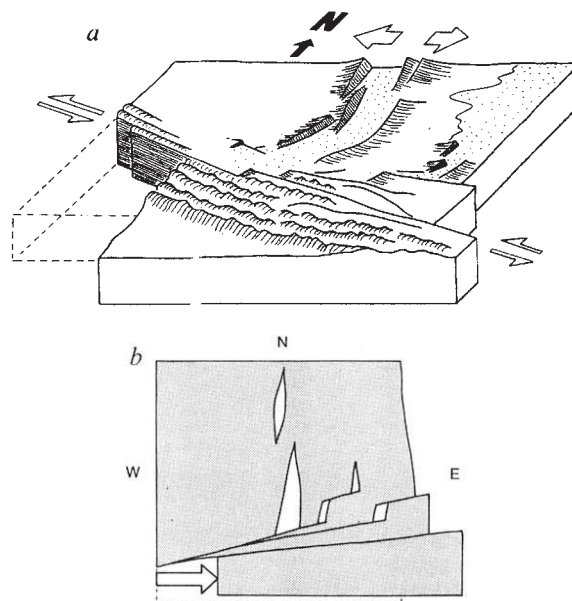


Fig. 6 *a*, Idealized block diagram showing how normal faults and grabens merge with strike-slip faults and absorb only part of the left-lateral displacements along them. *b*, Two-dimensional analogue: fissures and cracks branching off left-lateral strike-slip faults.

old. Note that, despite the extreme uncertainty about the erosion rates in the Qinling and ages of river channels of different orders within the drainage network, the order of magnitude of such rates is broadly comparable to that found along the Red River fault, another large strike-slip fault in China¹⁰, and is not inconsistent with the absence of a great earthquake along the Shang Xien fault in the past 1,000–2,000 yr (ref. 5). Third, the fact that the major branch of the Shang Xien-Dan Feng fault forms the southern boundary of the Shang Xien basin and the northern boundary of the Xi Ping basin (Fig. 1) suggests that these two basins, perhaps once part of a unique basin, have been displaced left-laterally by the transecting Shang Xien-Dan Feng fault. If this inference is correct, the total late Cenozoic offset of the Shang Xien-Dan Feng fault could be of the order of 150 km (Figs 1, 2).

Regional kinematics

The observations above show that the Neogene and Quaternary tectonics in the Qinling mountain range have been dominated by EW-ESE strike-slip faulting with minor components of normal faulting. In regions north of the Qinling, on the other hand (Figs 1, 2), extensional tectonics on NNE-ENE normal faults appears to have prevailed. Such a change in tectonic style corroborates the idea that the Fen Wei rift system (Shaanxi rift), as well as other normal faults, half graben and rifts north of the Qinling range may be viewed as analogous to fissures and cracks branching off left-lateral strike-slip faults, near their extremities (Fig. 6)^{6–9,12}. Because only part of the eastward movement of Tibet and Qaidam with respect to both north and south China is absorbed by thrusting and mountain building in the Qilian Shan and Longmen Shan, there remains a component of left-lateral movement between south and north China. Thus, as both Tibet and the Yangtze block are pushed towards the east with respect to north China by the northward movement of India, the strike-slip faults of the Qinling Shan actually represents splays of the eastern end of the Altyn Tagh-Haiyuan fault system. Although there are thrust components of slip along the Haiyuan fault in the Liupan Shan^{6,13}, the displacement between south and north China east of Baoji is absorbed not only along the Qinling faults but also by increasing amounts of NW-SE crustal extension, particularly in the grabens east of the Ordos block. The Qinling belt thus forms the southern limit of a deformed zone whose fragments are dragged eastward

by the Yangtze block, and the left-lateral slip observed along the Shang Xien and Lo Nan faults represents that part of the eastward displacement which is not taken up in the Fen Wei graben. The observation that these faults, or faults which splay off them, seem to continue along the edges of the Dabie Shan to vanish only $\approx 1,000$ km east of Xian in the region where they merge with the Tan Lu fault^{3,4,7}, suggest that gradual conversion, into extensional drag, of the remaining slip of south China relative to north China in the eastern Qinling, may have been the chief cause¹² of the late Cenozoic normal faulting observed in the north China basins¹⁴. Note that the fault pattern illustrated in Fig. 6, which involves predominantly normal faults coeval

with and merging at about 45° with predominantly strike-slip faults, and appears to be characteristic of intracontinental deformation^{15,16}, is observed at a smaller scale within the Quaternary sediments in the Wei He graben (Figs 1, 2, upper left).

This study was carried out within a Cooperative Research Project between the Chinese Academy of Geological Sciences, and Geological Bureau of Shaanxi province, and the French CNRS and INAG. We especially thank Li Tingdong, Xiao Xucheng, Li Chengda, Zhen Yli and G. Aubert for organization of the programme. M. Mattauer, P. Matte, J. Malavieille, Lu Yilun and Tang Yaoqing participated in the field studies. This is IGP contribution no. 828.

Received 29 April; accepted 21 August 1985.

1. Wang, H. S., Xu, C. G. & Zhou Zhengguo *Acta Acad. sin.* **56**, 270-280 (1982).
2. Chang, Q. S. *Metamorphic Geology of East Qinling* (People's Publishing House, Jiling, 1980).
3. Mattauer *et al.* *Nature* **317**, 496-500 (1985).
4. *East Qin Ling Geological map*, Scale 1:500,000 (Geological Regional Survey and Geological Bureau of Shaanxi province, 1969).
5. *Map of Historical Seismicity of China*, Scale 1:6,000,000 (Institute of Geophysics, Academia Sinica, 1976).
6. Tapponnier, P. & Molnar, P. *J. geophys. Res.* **82**, 2905-2930 (1977).
7. Molnar, P. & Tapponnier, P. *Science* **189**, 419-426 (1975).

8. Peltzer, G., Tapponnier, P. & Cobbold, P. *C.r. hebdom. Séanc. Acad. Sci., Paris* **294**, 1341-1348 (1982).
9. Tapponnier, P., Peltzer, G., Le Dain, A. Y., Armijo, R. & Cobbold, P. *Geology* **10**, 611-616 (1982).
10. Allen, C. R. *et al.* *Bull. geol. Soc. Am.* **95**, 686-700 (1984).
11. Berthé, D., Choukroune, P. & Jegouzo, P. *J. struct. Geol.* **1**, 31-42 (1979).
12. Peltzer, G. & Tapponnier, P. *J. geophys. Res.* (submitted).
13. Deng Qidong *et al.* *J. geophys. Res.* **89**, 4427-4445 (1984).
14. Ma Xingyuan, Deng Qidong, Wang Yipeng & Liu Hefu, *Z. Geomorph. N. F.* **42**, 99-116 (1982).
15. Tapponnier, P. & Molnar, P. *J. geophys. Res.* **84**, 3425-3459 (1979).
16. Armijo, R., Tapponnier, P., Mercier, J. L. & Han Tonglin, *J. geophys. Res.* (submitted).

LETTERS TO NATURE

Cosmological experiments in superfluid helium?

W. H. Zurek

Theoretical Astrophysics, Los Alamos National Laboratory,
Los Alamos, New Mexico 87545, USA

Symmetry breaking phase transitions occurring in the early Universe are expected to leave behind long-lived topologically stable structures such as monopoles, strings or domain walls¹⁻⁶. Here I discuss the analogy between cosmological strings and vortex lines in the superfluid, and suggest a cryogenic experiment which tests key elements of the cosmological scenario for string formation. In a superfluid obtained through a rapid pressure quench, the phase of the Bose condensate wavefunction—the ⁴He analogue of the broken symmetry of the field-theoretic vacuum—will be chosen randomly in domains of some characteristic size d . When the quench is performed in an annulus of circumference C the typical value of the phase mismatch around the loop will be $\sim (C/d)^{1/2}$. The resulting phase gradient can be sufficiently large to cause the superfluid to flow with a measurable (mm s^{-1}), randomly directed velocity.

Near the second-order phase transition, the potential contribution V to the free energy density of the system can be approximated by²⁻⁴:

$$V = \alpha |\psi(\vec{r})|^2 + \frac{1}{2} \beta |\psi(\vec{r})|^4 \quad (1)$$

Strings will be able to form when the order parameter (Higgs field) ψ is complex²⁻⁶.

$$\psi = |\psi| e^{i\theta} \quad (2)$$

In the broken symmetry phase α is negative and V defined above has the shape of a sombrero. This situation may occur in cosmologically relevant field theories¹⁻⁶. It is also a reasonable approximation for the superfluid ⁴He as well as for superconductors⁷⁻¹⁰. In particular, for the superfluid near critical temperature T_λ , the coefficient $\alpha = \alpha'(T - T_\lambda)$, where α' and β are constants which can be deduced from specific heat data^{7,8}.

Strings are axially symmetric, stationary solutions of the field theory with the potential given by equation (1). In the superfluid

⁴He the order parameter ψ can be thought of as a wavefunction of the Bose condensate^{8,9}. In Landau-Ginzburg theory it obeys the Gross-Pitaevskii equation⁷⁻⁹:

$$i\hbar\dot{\psi} = -(\hbar^2/2m)\nabla^2\psi - (\partial V/\partial|\psi|^2)\psi \quad (3)$$

When ψ is correctly normalized the mass m can be set equal to that of the ⁴He atom⁷⁻⁹. A string-like solution of the time-independent part of equation (3) can be expressed in cylindrical coordinates (ρ, ϕ, z) as: $\psi/\psi_0 = R(\rho/\xi) \exp(i n \phi)$ where $\psi_0 = \sqrt{\alpha/\beta}$ and ξ is the correlation length

$$\xi = \hbar/\sqrt{2m\alpha} \quad (4)$$

The radial part, $R(\rho/\xi)$, is zero at the origin, and remains small within a cylinder $\rho < \xi$. The winding number n in the azimuthal part of the solution must be an integer; otherwise the wavefunction ψ could not be single-valued. Analogous solutions of relativistic field theories are known as strings¹¹.

From the Schrödinger equation for ψ one can deduce that the velocity of the superfluid is given directly by the gradient of the phase θ of the wavefunction:

$$\vec{v}_s = (\hbar/m)\vec{\nabla}\theta \quad (5)$$

Thus the axially symmetric solution of the Gross-Pitaevskii equation predicts circulation of the superfluid around the axis of symmetry with the velocity:

$$v_\phi = (\hbar/m)n/r \quad (6)$$

Each atom of ⁴He involved in this circulation carries an amount $\hbar$ of angular momentum. The vorticity is therefore quantized.

Onsager¹² and Feynman¹³ proposed such quantized vortex lines to avoid a rotation paradox⁷⁻⁹ in the superfluid. They have been since thoroughly investigated in a variety of experiments¹⁴⁻¹⁶. They are usually created in the superfluid by stirring or by rotating. By contrast, the scenario (known as the 'Kibble mechanism') for the creation of their early Universe analogues—strings—in nonequilibrium phase transitions is thought to involve 'freezing out' of fluctuations of the Higgs field ψ in the course of a rapid phase transition induced by the decrease of temperature during the cosmological expansion¹⁻⁶. The key assumption states that the phase of the frozen-out field configuration will be determined independently in regions which did not have a chance to interact. By the same token, θ should differ in causally disconnected domains of the superfluid. In either

case, the density l of strings left behind by a rapid phase transition should be of the order of:

$$l \sim d^{-2} \quad (7)$$

Here d is the correlation length of the frozen-out configuration of the order parameter (Higgs field). The above formula follows from a natural assumption that each domain of size d has volume $\sim d^3$ and will contribute, on the average, a section of length d to a string^{4,5}.

Generation of vortex lines by the pressure quench in normal ⁴He has already been proposed¹⁷. However, the random network^{4,5} of vortex lines is likely to evolve rather rapidly and a quantitative verification of the above scenario may prove difficult. Here, we suggest an even more dramatic prediction relevant for the case of a pressure quench in an annular geometry. This experiment may be more challenging than generation of vortex lines¹⁷, but the expected effect—non-zero, measurable superfluid circulation—should remain unchanged for considerable periods of time which, in turn, should make quantitative comparisons with theory more reliable.

Consider rapid normal to superfluid phase transition occurring in an annulus with circumference $C = 2\pi R$, and with a minor radius $r \ll R$ (Fig. 1). Suppose, in accord with the Kibble mechanism, that the coherence length of the frozen-out configuration of ψ is d . The number of independent domains along the circumference of the annulus is:

$$N \cong C/d \quad (8)$$

If the phase of ψ is chosen independently in each of N domains, the typical mismatch of θ is given by:

$$\delta\theta = \int_C \vec{\nabla}\theta \cdot d\vec{s} \sim \sqrt{N} \quad (9)$$

where $d\vec{s}$ is the line element along the circumference of the annulus.

Therefore, the anticipated value of the gradient of θ 'frozen out' by the phase transition will be of the order of $\sqrt{N}/C$. We are forced to conclude that the phase transition will leave superfluid in an annulus circulating with a velocity of the order of:

$$v_s = (\hbar/m)/(Cd)^{1/2} \quad (10)$$

This formula is the key result of our paper. It amounts to a rather unexpected prediction: a rapid phase transition will leave superfluid circulating around an annulus.

To find out if v_s induced by pressure quench can be detected, we must estimate correlation length of fluctuations frozen out in the phase transition. To this end, consider liquid ⁴He near but above the phase transition temperature T_λ . The correlation length ξ of the order parameter and the correlation time τ determine spatial and temporal extent of individual fluctuations. Correlation time τ can be estimated from the Gross-Pitaevskii equation (3) as $\tau = \hbar/\alpha$. It can be also evaluated directly by dividing correlation length by the characteristic velocity with which fluctuations propagate. In our case, this characteristic velocity is of the order of the velocity of the second sound, c_2 , in the superfluid. Let us first suppose that the phase transition is accomplished on a very fast ($\ll \tau$) timescale. Critical slowing down and the fact that the velocity of the first sound, c_1 , near T_λ is much larger than c_2 (refs 7, 9), make pressure changes on a timescale faster than τ feasible. Such rapid phase transition will be completed before the initial configuration of ψ has a chance to adjust. Hence, in this case of an instantaneous quench the size of the 'frozen-out' fluctuation is simply equal to the initial, pretransition correlation length ξ_i , that is, $d = \xi_i$. One can now use equation (10) to calculate circulation velocity v_s .

This simple estimate of d illustrates the connection of the initial configuration of ψ with the characteristic scale d , but it also avoids a central question: what scales of d can one hope to achieve by performing a quench on a realistic, and hence finite timescale? For, in contrast to the case considered above, had one performed the transition infinitesimally slowly, correla-

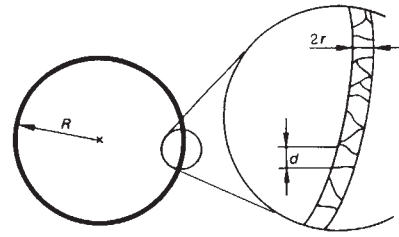


Fig. 1 In an annulus with circumference $C = 2\pi R$ quench will result in $\sim C/d$ 'domains' of average size d with significantly different phase θ of the Bose condensate wavefunction ψ . The typical mismatch of the 'phase around the loop', $\delta\theta = \int_C \vec{\nabla}\theta \cdot d\vec{s} \sim \sqrt{C/d}$, corresponds to the measurable superfluid velocity $v_s \sim (\hbar/m)/\sqrt{Cd}$.

tion length would have had time to become infinite at $T = T_\lambda$, and $d \gg C$ would have implied $v_s = 0$. Clearly, the intermediate case of finite speed quench is most realistic. To discuss it, let us suppose that the temperature parameter ϵ changes linearly with time:

$$\epsilon = (T_\lambda - T)/T_\lambda = t/\tau_Q \quad (11)$$

Equilibrium correlation length would then vary as:

$$\xi = \xi_0 |\epsilon|^{-\nu} \quad (12)$$

where $\xi_0 = 5.6 \text{ \AA}$ and $\nu = \frac{1}{2}$ for the Landau-Ginzburg theory, while the renormalization group as well as experiments yield $\nu = 2/3$ and $\xi_0 = 4 \text{ \AA}$ (ref. 16). Moreover, the characteristic velocity can be expressed as:

$$u = u_0 \epsilon^{1-\nu} \quad (13)$$

with $u_0 = 70 \text{ m s}^{-1}$ in Landau-Ginzburg theory and $u_0 = 47 \text{ m s}^{-1}$ in the renormalization group case¹⁶. The correlation timescale τ can be now estimated as:

$$\tau = \xi/u = \tau_0/\epsilon \quad (14)$$

If changes of ϵ occur on a timescale small compared to τ , the state of the system—fluctuations of ψ —can readjust to the new set of thermodynamic parameters, and the system can be regarded as evolving close to equilibrium. On the other hand, if τ exceeds the inverse of the rate of change of ϵ , ψ will lag and the situation will be close to the instantaneous phase transition we have discussed previously. The freeze-out time $\hat{t}$ which marks the boundary between these two regimes occurs when τ is equal to the time interval before the transition (Fig. 2). This implies the equality:

$$\tau(\hat{t}) = \hat{t} \quad (15)$$

Using equation (14) we calculate the freeze-out time:

$$\hat{t} = \sqrt{\tau_0 \tau_Q} \quad (16)$$

This readily yields out estimates of the frozen-out correlation length:

$$d(\hat{t}) = \xi_0 (\tau_Q/\tau_0)^{\nu/2} \quad (17)$$

Above, $\tau_0 = \xi_0/u_0 \cong 8.5 \times 10^{-12} \text{ s}$ in Landau-Ginzburg as well as in the renormalization group models. It should be emphasized that equation (17) is only an estimate: it can be expected to yield values correct to within an order of magnitude. Moreover, as d corresponds to the size of the causal (sonic) horizon at the freeze-out time, it is probably an overestimate. We can, nevertheless, use it in equation (10) to obtain an estimate of the circulation velocity v_s induced by the rapid quench.

To this end, we set conservatively $\tau_Q = 10^{-2} \text{ s}$ and $C = 1 \text{ cm}$. This yields $d \sim 10^4 \text{ \AA}$ and a detectable $v_s \sim 1 \text{ mm s}^{-1}$. As the values mentioned above appear within the reach of experimental possibilities and as the final velocities are significant, we are also forced to consider the issue of conservation of the angular momentum. To formulate it more conveniently, we make one more simplifying assumption about the design of the experiment, by taking $d \sim r \gg \xi_f$ (ξ_f , final correlation length). This construc-

tion ascertains that it will be the overall phase around C rather than creation of individual vortices inside the annulus which will determine final values of v_s . It also assures that boundary effect can be neglected after the quench is completed.

The typical angular momentum carried by the superfluid:

$$J_s = MRv_s \quad (18)$$

can be expected to be of the order of $5 \times 10^{-11} \text{ g cm}^2 \text{ s}^{-1}$ for

$$M = \rho_s C \pi r^2$$

where we take typical $\rho_s \sim 0.1 \text{ g cm}^{-3}$ as the density of the superfluid, and, as before, $C = 1 \text{ cm}$ and $r \sim d = 10^{-4} \text{ cm}$. Brownian motion of the superfluid is compatible with

$$J_T = (kTMR^2)^{1/2} \quad (19)$$

which, at $T \sim T_\lambda \sim 2 \text{ K}$, is $\sim 10^{-13} \text{ g cm}^2 \text{ s}^{-1}$. There is a discrepancy of several orders of magnitude between the anticipated value of J_s and the much smaller thermal J_T . One may wonder whether the scenario which appears to predict such apparent violations of conservation laws in the context of superfluid phase transitions can be regarded as a possible model for generation of cosmological defects. While these questions will have to be reconsidered after experimental results become available, it is of interest to anticipate possible answers.

Let us first suppose that velocities induced by the rapid phase transition are consistently smaller than v_s predicted by equation (10) and that final angular momenta are compatible with J_T . This would indicate that either the scenario for the formation of topological defects has a flaw, or that for some reason the cosmological scenario is not applicable to ^4He . It is perhaps worth noting that such an outcome ($v_s \sim 0$) would not violate causality. Indeed, there is an analogous situation—non-separable Einstein-Podolsky-Rosen correlations of two quantum systems—in which individual subsystems appear to be acting in a completely random fashion, and yet correlations between them are sufficiently strong to ensure conservation of angular momentum¹⁸. It is conceivable that a similar situation could arise for fluctuations of ψ : they could appear to be independent and random on a scale of a few correlation lengths but the conservation law could at the same time restrict the 'phase around the loop' to much less than the value predicted by equation (9). If one were forced by the results of the experiment to invoke this explanation for ^4He , standard cosmological scenarios for monopole, string and domain formation would have to be reconsidered.

I regard the outcome outlined above to be rather unlikely. Let me therefore suppose that velocities detected in the quench experiments will be consistent with those predicted by equation (10). How can one reconcile these with the law of conservation of angular momentum? We start by noting that there is a simple condition which must be satisfied if the superfluid velocity is to equal the velocity of brownian motion. From equations (18) and (19) it follows that $J_s = J_T$ when:

$$\rho_s (\pi r^2 / d) = (m / \hbar)^2 kT \quad (20)$$

Clearly, this condition cannot be satisfied far away from the λ -line, where $\rho_s \sim 0.1 \text{ g cm}^{-3}$. However, superfluid density is a function of temperature¹⁶:

$$\rho = \rho_0 \epsilon^{0.67} \quad (21)$$

with $\rho_0 = 0.36 \text{ g cm}^{-3}$. Therefore close to T_λ the superfluid density will become sufficiently small for equation (20) to hold. If we now set $r = ad$, $a \sim 1$, to ensure that boundary conditions do not matter at a time t when the flow is supposed to start, we will discover that equation (20) holds when ρ_s is also calculated at ϵ corresponding to t . Indeed, equation (20) can then be written as

$$\rho_s d = (m^2 / \hbar^2) kT / \pi a^2 \quad (22)$$

and its left-hand side equals, for the data quoted by Ahlers¹⁶, $\rho_s d = 1.44 \times 10^{-8} \text{ g cm}^{-2}$ while the right-hand side yields $(m^2 / \hbar^2) kT_\lambda / \pi a^2 = 0.35 \times 10^{-8} / a^2 \text{ g cm}^{-2}$.

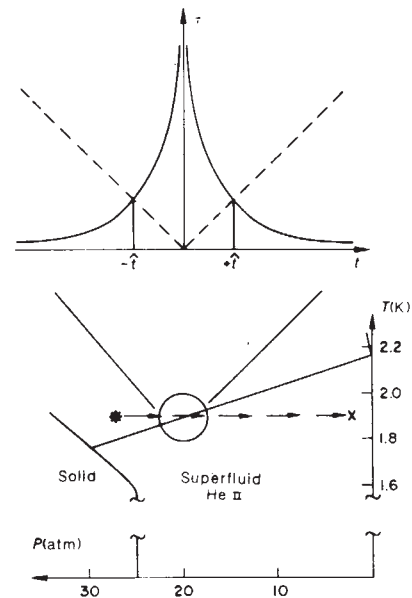


Fig. 2 Changes of the fluctuation timescale during a quench; a portion of the phase diagram of ^4He is shown in the lower part of the figure. The approximate trajectory of the quench is also indicated. Near the λ -line timescale τ becomes large, and fluctuations of ψ 'freeze out'. Characteristic freeze-out time $\hat{t}$ determines the correlation length of the configuration responsible for creation of topological defects, as well as for the remanent circulation in annular geometry.

The two sides of this equality, equation (20), are sufficiently close to each other that setting $r \sim d/2$ (diameter equal to the correlation length—a fairly natural assumption) can make them numerically almost identical. Clearly, at the beginning of the superflow, at $t = \hat{t}$ after the phase transition, angular momentum carried by the superfluid is due to the brownian motion.

The resolution of the paradox is now at hand; it is a well-known experimental fact that in the course of changes of ϵ the velocity of the superfluid remains constant, and its angular momentum changes accordingly due to the temperature-dependent ρ_s (ref. 19). The violation of the conservation law can be, nevertheless, avoided, as the normal fluid takes up angular momentum and transfers it to the walls of the container. Thus, as ϵ increases during the quench, normal fluid will begin to counter flow with the appropriate velocity, and any conflict with conservation laws can be avoided. We are, however, left with a curious equality which ought to be—and appears to be—satisfied. Indeed, equation (22) has been proposed^{16,20,21} in discussions of the healing length of superfluid helium. We have arrived at it in a very different way.

We have proposed an experiment—quench-induced phase transition from normal ^4He to the superfluid—which, when analysed using the standard scenario for formation of topological defects in phase transitions, predicts: (1) appearance of vortex lines in bulk superfluids, and (2) measurable random velocities of circulation in torus-shaped containers. The final velocity of circulation (mm s^{-1}) is measurable. This experiment, therefore, can be regarded as a laboratory test of the Kibble mechanism, one of the key elements of the cosmological scenario.

I thank Dr Laurence Campbell, Professor Richard Feynman, Dr James Hoffer, Professor John Wheatley and Ms Shouu-Shinn Shiah for useful conversations. This research was supported by the J. Robert Oppenheimer Fellowship in Theoretical Astrophysics at the Los Alamos National Laboratory.

Received 28 June; accepted 2 August 1985.

1. Zel'dovich, Ya. B., Kobzarev, I. Yu. & Okun, L. B. *Zh. eksp. teor. Fiz.* **67**, 3–11 (1974); *Soviet Phys. JETP* **67**, 401–409 (1975).
2. Kibble, T. W. B. *J. Phys. A*, **9**, 1387–1398 (1976).
3. Linde, A. D. *Rep. Prog. Phys.* **42**, 389–437 (1979).
4. Vilenkin, A. *Phys. Rev. D* **24**, 2082–2089 (1981).

5. Albrecht, A. & Turok, N. *Phys. Rev. Lett.* **54**, 1868-1871 (1985).
6. Kibble, T. W. B. *Phys. Rep.* **67**, 183-199 (1980).
7. Ginzburg, V. L. & Pitaevskii, L. P. *Zh. eksp. teor. Fiz.* **34**, 1240-1247 (1958); *Soviet Phys. JETP* **34**, 858-863 (1958).
8. Lifshits, E. M. & Pitaevskii, L. P. *Landau and Lifshits Course in Theoretical Physics*, Section III (Pergamon, Oxford, 1980).
9. Tilley, D. R. & Tilley, J. *Superfluidity and Superconductivity* (Wiley, New York, 1974).
10. Aitchison, I. J. R. & Hey, A. J. G. *Gauge Theories in Particle Physics* (Hilger, Bristol, 1982).
11. Nielsen, H. B. & Olesen, P. *Nucl. Phys. B* **61**, 45-55 (1973).
12. Onsager, L. *Nuovo Cim.* **6**, Supp. 2, 249 (1949).
13. Feynman, R. P. in *Progress in Low Temperature Physics* (ed. Gorter, C. J.) Vol. 1, Ch. II (North Holland, Amsterdam, 1956).
14. Vinen, W. F. *Proc. R. Soc. A* **240**, 114-128 (1957).
15. Vinen, W. F. *Proc. R. Soc. A* **260**, 218-230 (1961).
16. Ahlers, G. in *The Physics of Liquid and Solid Helium* Vol. 2, (eds Benneemann, K. H. & Ketterson, J. B.) Ch. II (Wiley, New York, 1976).
17. Zurek, W. H. in *Proc. a. Meet. Particle and Field Theory Division of the Am. phys. Soc., Santa Fe* (eds Goldman, T. & Nieto, M. M.) 479 (World Scientific, Singapore, 1985).
18. Wheeler, J. A. & Zurek, W. H. (eds) *Quantum Theory and Measurement* Ch. III (Princeton University Press, 1983).
19. Reppy, J. D. & Depatie, D. *Phys. Rev. Lett.* **12**, 187-189 (1964).
20. Ferrell, R. A., Menyhard, N., Schmidt, H., Schwabl, F. & Szeftalussy, Ann. *Phys.* **47**, 565-613 (1961).
21. Halperin, B. I. & Hohenberg, P. C. *Phys. Rev.* **177**, 952-971 (1969).

Does Saturn have rings outside $10 R_s$?

A. F. Cheng

Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland 20707, USA

L. J. Lanzerotti & C. G. MacLennan

AT & T Bell Laboratories, Murray Hill, New Jersey 07974, USA

Vasundhara *et al.*¹ and Mahra *et al.*² have recently reported detections of new saturnian ring systems near 12.5 and 19.5 Saturn radii (R_s), respectively. Lazarus *et al.*³ have previously suggested that one or more tenuous rings may exist between 14 and $19 R_s$, based on observations of unexplained drops in low-energy plasma ion densities in this region of the magnetosphere by Pioneer 11 and Voyagers 1 and 2. However, Baron and Elliot⁴ have placed a stringent upper limit on the I -band brightness of any material in the region 10 – $36 R_s$ west of Saturn. Here we review the Voyager ion and electron data in the energy range 30 – $1,000$ keV as measured by the Low Energy Charged Particle (LECP) experiment. There is no convincing evidence for ring matter outside $10 R_s$ in these data. Features in the charged particle fluxes in these regions are more readily explained by temporal variations and/or spatial structure unrelated to ring matter, such as the mantle on the dayside⁵ and/or detached plasma sheets⁶.

Vasundhara *et al.*¹ observed the occultation of the star SAO 158913 from two stations, Kavalur and Naini Tal. Kavalur observed both immersion and emersion events near $12.5 R_s$, but Naini Tal observed only an immersion due to 'cloudy sky conditions'. Examination of the light curves shown in Fig. 2 of ref. 1 reveals that the immersion and emersion events at Kavalur clearly differ from each other and are qualitatively different from the immersion event at Naini Tal. It is perhaps even more significant that Vasundhara *et al.*¹ do not report any intensity variations near $19.5 R_s$. Likewise, Mahra *et al.*² observed the occultation of SAO 158763 from Naini Tal, apparently using the same techniques as Vasundhara *et al.*¹. An immersion event was reported near $19.5 R_s$, with sharp intensity decreases exceeding 0.3 mag, but no emersion event was observed there, within the observational scatter of 0.03 mag. Furthermore, no events are reported near $12.5 R_s$.

Significant discrepancies therefore exist among the occultation observations, arguing against the presence of ring matter at either location. An additional constraint is provided by the observations of Baron and Elliot⁴, who showed that the I magnitude was fainter than $+22$ arc s⁻² for material distributed over a region larger than the 3 arc s seeing disk. We assume the colour index $V - I = 0.81$ and the apparent visual magnitude of the sun

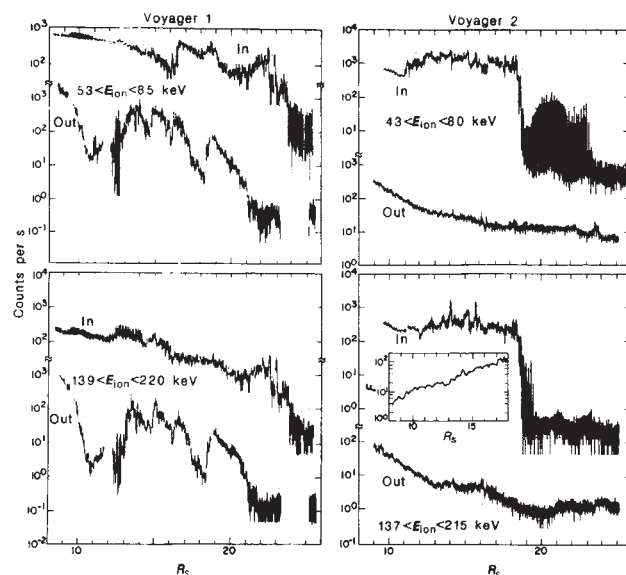


Fig. 1 Ion count rates in two energy channels from the LECP experiments during the encounters of the Voyager 1 and 2 spacecraft with the Saturn system. The inset shows ion phase space densities for Voyager 2 inbound.

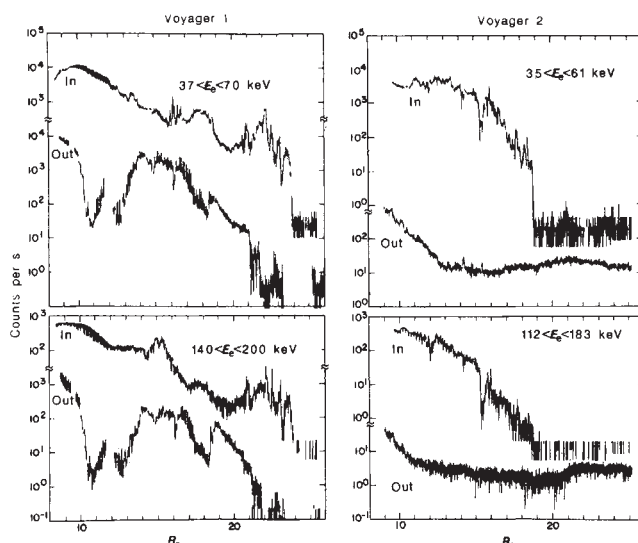


Fig. 2 Electron count rates in two energy channels from the LECP experiments during the encounters of the Voyager 1 and 2 spacecraft with the Saturn system.

$V_0 = -26.74$, as usual. The occultation measurements^{1,2} suggest an average optical depth $\tau \approx 0.1$. If this optical depth applies throughout the seeing disk, then the obscuring matter must have an extremely low albedo. Specifically, if $V > 22.81$ arc s⁻², the product of geometric albedo p and phase function ϕ must be $p\phi < 2 \times 10^{-6}$. Neglecting ϕ , we find an albedo $\sim 10^4$ times lower than that of any known body in the Solar System. Even the dark regions of Iapetus have $p \sim 0.05$. The occultation measurements^{1,2} suggest a radial extent for the rings of 0.2 – $0.5 R_s$, which is large enough to fill approximately the seeing disk reported by Baron and Elliot⁴, as was assumed above.

Energetic ion and electron data from the Voyager 1 and 2 LECP experiments are shown in Figs 1 and 2, respectively. The data inbound and outbound are ordered by distance from the planet. Inbound, the magnetopause was crossed by Voyager 2 near the noon meridian between 18 and $19 R_s$. Outbound, on the morning side, the magnetopause was at $\sim 50 R_s$. For Voyager 1, the last inbound magnetopause crossing occurred at $\sim 23 R_s$, while the first outbound crossing was at $\sim 43 R_s$. If long-lived ring systems are present, absorption of charged particles by ring

matter should produce an observable effect such as a gradient change in the count rate profiles. No such effects are evident in the data or in the Voyager 2 inbound ion-phase space densities (inset in Fig. 1). The depression of the Voyager 2 electron data inbound at $\sim 15 R_s$ would correspond better with a 'moon' signature. The 'corresponding plateau' in the ion data (also at $\sim 15 R_s$) has the wrong sign for ring absorption.

Unfortunately, the interpretation of the charged particle data is complicated by many such competing effects. A 'plasma mantle' has already been identified at $\sim 11 R_s$ on Voyager 2 inbound and $\sim 12.5 R_s$ on Voyager 1 inbound⁵. Inside the mantle boundary, the count rate profiles in the low energy ion and electron channels are suggestive of 'permanently' trapped particles, whereas outside it, to the magnetopause, the profiles suggest a quasi-trapped population. This mantle boundary clearly complicates the search for ring matter near $12.5 R_s$. Furthermore, Voyager 2 inbound encountered the magnetopause near $19 R_s$, as determined by plasma, magnetic field, high energy charged particle, and plasma wave data. On Voyager 1 outbound, an unexplained drop-out occurred at all observed ion and electron energies from 10 to $14 R_s$ (ref. 5). No such effect has been observed on any other pass.

Thus the structure and dynamics of the saturnian magnetosphere must be taken into account in interpretations of charged particle data. We have therefore identified the Voyager 2 outbound pass as being especially well suited for detection of ring matter, since the magnetosphere was unusually quiescent and spatially distended during that time. The magnetopause was not encountered until $\sim 50 R_s$ outbound, and the mantle boundary

was not encountered within $25 R_s$. Temporal variations also appeared to be minimal during that pass. Thus, the Voyager 2 outbound pass is the most likely to reveal the effects of ring matter.

To qualify as a signature of ring matter, an absorption feature must be observed consistently at the same radius in many different ion and electron channels, both at high energies (for example, lower panels of the Voyager 2 portions of Figs 1 and 2) and at low energies. Even for the Voyager 2 outbound pass, the data in Figs 1 and 2 do not reveal any such absorption signatures.

We conclude that the reported detections of saturnian ring systems near $12.5 R_s$ and $19.5 R_s$ are subject to serious question. Significant discrepancies exist among the various reported occultation observations. If the rings actually exist, they must be composed of matter possessing a spectacularly low albedo, roughly four orders of magnitude less than carbon black. Finally, no convincing evidence for ring matter outside $10 R_s$ is found in Voyager energetic charged particle data.

We thank Dr M. Paonessa for calculating the phase space densities. The work at JHU/APL was partially supported by NASA under contract N00024-83-C-5301.

Received 25 March; accepted 12 August 1985.

1. Vasundhara, R. *et al.* *Nature* **312**, 621–623 (1984).
2. Mahra, H. S., Pandey, A. K., Mohan, V. & Sanwal, B. B. *Nature* **313**, 38–39 (1985).
3. Lazarus, A. J., Hasegawa, T. & Bagenal, F. *Nature* **302**, 230–232 (1983).
4. Baron, R. L. & Elliot, J. L. *Astr. J.* **88**, 562–564 (1983).
5. Krimigis, S. M. *et al.* *J. geophys. Res.* **88**, 8871–8893 (1983).
6. Goertz, C. K. *Geophys. Res. Lett.* **10**, 455–458 (1983).

Are all the 'martian' meteorites from Mars?

U. Ott & F. Begemann

Max-Planck-Institut für Chemie, Otto-Hahn-Institut,
6500 Mainz, FRG

The shergottites, nakhlites and Chassigny, collectively known as the SNC meteorites, seem to define a petrological and geochemical sequence¹. Together with their young radiometric solidification ages of only $\sim 1,300$ Myr (refs 2–9), this has been interpreted as evidence for magmatic activity on their parent body late in its history. Such late magmatism is most easily envisaged for planetary-sized bodies and it has been suggested that SNC meteorites are derived from Mars (see ref. 10 and refs therein). This suggestion is supported by the apparent similarities between the noble gas and nitrogen composition of the present martian atmosphere^{11,12} and gases trapped in shock-produced glass phases of the shergottite EETA 79001 (refs 13, 14). Our analyses of trapped argon, krypton and xenon in Shergotty, Nakhla and Chassigny reported here show, however, that these meteorites do not contain unadulterated noble gases from the martian atmosphere, fractionated or not. They indicate that for Nakhla and Chassigny, in particular, a martian origin is only possible if the martian atmosphere contains a higher fraction of radiogenic ^{129}Xe and of fissionogenic xenon than does xenon in martian rocks. This is the opposite of the situation on Earth.

The gas concentrations as well as elemental abundance ratios in Shergotty, Nakhla and Chassigny are very different from one another and from those in EETA 79001 and the martian atmosphere (Table 1). This by itself argues neither against a common origin of the meteorites nor against their origin from Mars. For the gas patterns of the meteorites to be similar to one another, the mode of occurrence of the gases and of their acquisition by the meteorites must be similar. In the particular case of a resemblance of the meteoritic gases with those in the martian atmosphere, a process must be found to introduce gases into rocks without much fractionation. Shock appears to be such a process. The resemblance between the patterns of EETA 79001

glass and the atmosphere of Mars is, therefore, generally ascribed to the introduction by shock of atmospheric gases into the glassy constituents on their conversion into the glassy state. That the gas concentrations per cm^3 in both are within a factor of 2 suggests that this conversion has taken place at the surface of Mars or at least by direct contact between what is now meteorite EETA 79001 and the atmosphere. Gases from pore space in some deeper layers on Mars would presumably not be sufficient.

Only the glass in EETA 79001 exhibits these features. With no such gas-rich glass reported for Shergotty, Nakhla and Chassigny, these meteorites must have acquired their primordial noble gases in some other way. Depending on the exact kind of process(es), this could cause fractionation between the gases, but it should not yield perceptible isotope fractionation. Because in Shergotty, Nakhla and Chassigny at least one isotope ratio is different from the corresponding one in EETA 79001 glass or the martian atmosphere it is clear that these meteorites do not contain unadulterated noble gases from the present-day martian atmosphere, fractionated or not.

A more difficult question is whether the gases are a mixture of two or more components, with one component being the martian atmosphere. The plots of noble gas data in Fig. 1a, b have the property that any mixture between two gas components falls on the straight line connecting the pure endmembers. In the $^{129}\text{Xe}/^{132}\text{Xe}$ versus $^{36}\text{Ar}/^{132}\text{Xe}$ plot (Fig. 1a), Chassigny, Shergotty and EETA 79001 glass do form a linear array, with Shergotty being intermediate. Hence, its gases may be a mixture of Chassigny and EETA 79001 glass-type gases. The same conclusion is reached from the $^{129}\text{Xe}/^{132}\text{Xe}$ versus $^{84}\text{Kr}/^{132}\text{Xe}$ plot (Fig. 1b), where again Shergotty plots on the line between Chassigny and EETA 79001 glass, and from the ratio of ^{40}Ar to trapped ^{36}Ar . In five separate analyses, this ratio in the gases released at 800°C , $1,200^\circ\text{C}$ and $1,600^\circ\text{C}$ was remarkably constant (2,265–2,580) and indistinguishable from that of $\sim 2,500$ reported for EETA 79001 glass¹⁴ or the $3,000 \pm 1,000$ obtained during the Viking mission¹¹.

Figure 1b also shows that Nakhla falls off and above the mixing line. If Nakhla is to contain a mixture of EETA 79001 glass-type and Chassigny-type gases, a third component must be present. This may be pure radiogenic ^{129}Xe from the decay

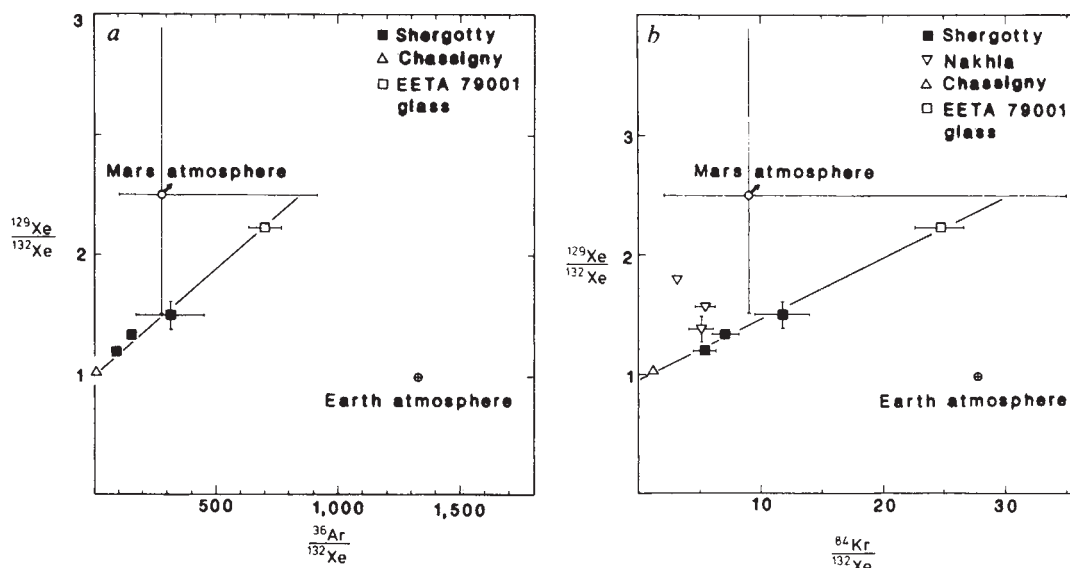


Fig. 1 Isotopic ratio $^{129}\text{Xe}/^{132}\text{Xe}$ versus $^{36}\text{Ar}_{\text{trapped}}/^{132}\text{Xe}$ (a) and versus $^{84}\text{Kr}/^{132}\text{Xe}$ (b). Shergotty data points fall on a mixing line between Chassigny-type and EETA 79001 glass-type¹⁴ gases. Data points for Nakhla plot off this line in b. No data points for Nakhla are shown in a because of the predominance of cosmogenic ^{36}Ar in this meteorite.

of now extinct ^{129}I , which decays with a mean lifetime of 25 Myr into ^{129}Xe .

Such a decomposition of Nakhla gases into hypothetical components is not unique, the ambiguity being due, in part, to difficulties of accurately deriving the amount of ^{36}Ar trapped in Nakhla. This is because the high cosmic-ray exposure age of 10 Myr (refs. 15 and work in preparation) together with a high Ca content of ~11% (ref. 16) make cosmogenic ^{36}Ar predominant (Table 1). Hence, no additional information can be derived from the presence or absence of a correlation in the $^{129}\text{Xe}/^{132}\text{Xe}$ versus $^{36}\text{Ar}/^{132}\text{Xe}$ plot of Fig. 1a. However, this should not detract from the fact that the noble gas pattern in Nakhla cannot be generated by any mixture of the gases from Chassigny and EETA 79001 glass.

The deviation from the mixing line may be even larger than is apparent from Fig. 1b because the highest $^{129}\text{Xe}/^{132}\text{Xe}$ ratio of 1.79 is presumably closest to the true ratio in Nakhla, the lower ones possibly being caused by different admixtures of terrestrial atmospheric xenon with $^{129}\text{Xe}/^{132}\text{Xe} = 0.983$. The

alternative explanation would be that different 100-mg samples of the meteorite contain xenon of different $^{129}\text{Xe}/^{132}\text{Xe}$ ratios. Although this is frequently encountered in primitive meteorites where iodine-rich phases still contain evidence of the former existence of ^{129}I , we find it hardly compatible with a magmatic history of the meteorite if this is to have lasted until, or have been resumed again shortly before, 1,300 Myr ago.

The essential distinction between the noble gases in the two endmembers of the mixing line, Chassigny and EETA 79001 glass, lies in different isotope abundance ratios. Such a difference exists not only for the relative abundance of ^{129}Xe but extends to the heaviest isotopes of xenon (Fig. 2). Here, again, the relative abundances are lower in Chassigny than in EETA 79001 glass. Such variations are generally ascribed to the presence, in varying proportions, of xenon from the decay of ^{244}Pu or ^{238}U . The data shown in Fig. 2 indicate that, relative to EETA 79001 glass, there is a deficit in Chassigny of both radiogenic ^{129}Xe and of fissionogenic xenon.

The differences between the noble gas fingerprints of

Table 1 Noble gases in 100–700-mg meteorite samples after a preheat step of 30 min at 800 °C (Chassigny 30 min at 1,200 °C)

	Extraction temperature (°C)	$^{36}\text{Ar}_{\text{spall.}}$	$^{36}\text{Ar}_i$	$^{36}\text{Ar}_i/^{132}\text{Xe}$	$^{84}\text{Kr}/^{132}\text{Xe}$	$^{129}\text{Xe}/^{132}\text{Xe}$
Shergotty						
KP 3	1,600	18.8 ±2.2	8.17 ±2.15	90 ±25	5.4 ±0.9	1.186 ±0.036
SI/III/IV	1,200	9.81 ±1.16	7.42 ±1.18	156 ±29	7.1 ±1.1	1.338 ±0.028
	1,600	8.18 ±0.96	2.09 ±0.91	312 ±143	11.9 ±2.3	1.499 ±0.112
Nakhla						
P1	1,600	115 ±11	7.6 ±10.1	—	5.1 ±1.1	1.378 ±0.114
H1	1,600	128 ±13	7.7 ±11.2	—	3.1 ±0.5	1.793 ±0.056
P 2/3/4	1,600	107 ±11	7.3 ±9.4	—	5.7 ±0.9	1.557 ±0.031
Chassigny						
PB 4	1,600	12.4 ±1.8	2.11 ±1.66	9.2 ±7.3	1.1 ±0.1	1.029 ±0.019
EETA 79001 glass	Total	—	172 ±12	700 ±70	24.6 ±2.0	2.231 ±0.037
Mars atmosphere (range)	—	—	410	280 (100–900)	8.8 (2–35)	2.5 (1.5–4.5)

The full data and experimental details will be reported elsewhere. Gas concentrations are in units of $10^{-10} \text{ cm}^3 \text{ STP per gram}$ except for the martian atmosphere where it is per cm^3 (equivalent to partial pressure, in atmospheres, of ^{36}Ar at the surface of Mars). Mars data are from refs 11, 12, EETA 79001 from ref. 14. Ar_i , trapped argon.

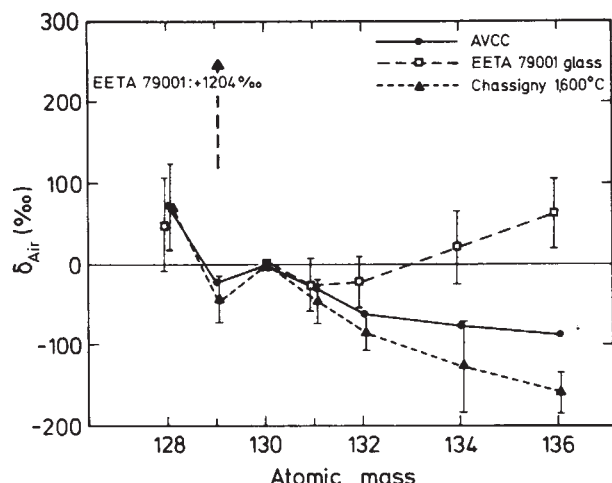


Fig. 2 Isotopic composition of xenon in EETA 79001 glass¹⁴ and Chassigny, shown as deviation from the composition of terrestrial atmospheric xenon. Normalization is to ^{130}Xe . Low abundant ^{124}Xe and ^{126}Xe are not shown because in Chassigny contributions from cosmogenic xenon are non-negligible. The other isotopes have been corrected on the basis of the ^{126}Xe abundance. Raw data in δ -values are:

$$124/126/128/129/130/131/132/134/136$$

$$= (+230 \pm 152)/(+525 \pm 464)/(+93 \pm 44)/(-50 \pm 26)/$$

$$= 0/(-47 \pm 25)/(-93 \pm 19)/(-136 \pm 55)/(-167 \pm 23).$$

The composition of average carbonaceous chondrite (AVCC)-xenon (frequently encountered in primitive meteorites) is shown for comparison.

Shergotty, Nakhla, Chassigny and EETA 79001 glass do not rule out a common parent body. This is evident from terrestrial rocks where similar variations of the elemental and isotope abundance ratios do occur (see ref. 17). It seems that on Earth, even after 4,500 Myr of thermal activity, there are still reservoirs which have not exchanged or equilibrated their noble gases. This is apparently true not only for radiogenic species such as ^4He , ^{40}Ar , and fissionogenic xenon which are still being generated, but also for radiogenic ^{129}Xe which has been present for more than 4,000 Myr. Note, however, that whenever variations in isotope abundance ratios are found for terrestrial samples, they are always in the sense that the rocks contain a higher proportion of radiogenic nuclides than does the atmosphere. This is expected for any closed-system model of uniform degassing of a planet, be it early or late, catastrophic or continuous degassing. Hence, if the noble gases of the martian atmosphere are represented by the gases in EETA 79001 glass (in particular if the isotopic composition of xenon in the martian atmosphere is like that in EETA 79001 glass) and if Shergotty, Nakhla and Chassigny are also to come from Mars, then the situation on Mars must be the opposite of that on Earth in that the atmosphere would have to contain a higher proportion of radiogenic xenon isotopes than does solid Mars. By inference, the development of the martian atmosphere must have been quite different from that of the terrestrial atmosphere.

Among the possible explanations we can envisage is, first, preferential release of radiogenic xenon into the martian atmosphere. Stepwise heating of primitive meteorites shows that radiogenic ^{129}Xe and fissionogenic xenon are often less retentively sited than trapped xenon (see refs 18, 19). However, for the retentivity of the original sites of the xenon components to be relevant, the degassing must occur before the starting material is completely melted. Thus, the source rocks of xenon in the martian atmosphere must either be a partial melt or, if they are or ever were completely molten, the degassing must have been much less effective from the melt than it was from the partial melt. Considering the large increase in diffusion length on going from

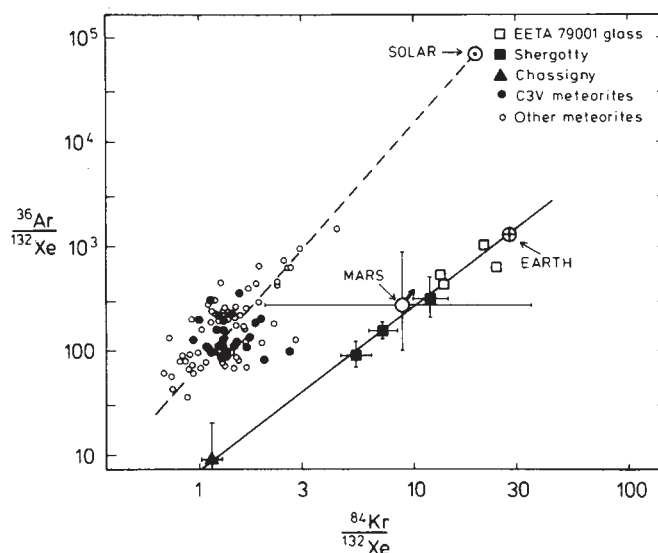


Fig. 3 Ratio of trapped $^{36}\text{Ar}/^{132}\text{Xe}$ versus $^{84}\text{Kr}/^{132}\text{Xe}$ on a log-log scale. Note that most meteorites, including the C3V meteorites, which have been suggested as veneer material supplying Mars with its volatiles²⁰, scatter around a line which may pass through the solar point. SNC meteorites define a different linear array passing through the data points for the atmospheres of Mars^{11,12} and Earth.

partial to total melt, this does not seem to be impossible particularly since it is not known whether convection in the molten state on Mars is as effective as it seems to be on Earth.

As to the time of the putative degassing, this must have happened until or after the decay of ^{129}I and until or after most of the fissionogenic xenon had been produced. If the parent nuclide was ^{244}Pu (half-life of 83 Myr) a few hundred million years would suffice for most of it to decay, but if the source was ^{238}U , the degassing would have to have happened very late in the martian history.

Second, one could envisage a late addition to Mars of a thin veneer of matter which brought in essentially all the noble gases which are in the atmosphere at present. This has already been proposed by Anders and Owen²⁰, who argue that, per unit mass of the planet, the martian atmosphere contains 100–1,000 times less primordial noble gases than is contained in meteorites which are presumably close analogues of the building blocks of Mars. An addition of 1% or so of such meteoritic matter, if completely degassed, would, therefore, suffice to account for the present martian atmosphere. (Incomplete degassing of this veneer with preferential release of ^{129}Xe and fission xenon into the atmosphere amounts to our first alternative.) A problem with this proposal is that only a sub-group of the carbonaceous chondrites (the C3Vs) may have a high enough $^{129}\text{Xe}/^{132}\text{Xe}$ ratio to qualify as veneer material—and then only with difficulties. To this we must now add the implication that, in this model, the noble gases in martian rocks and in the martian atmosphere would be essentially unrelated and independent of one another.

Figure 3 shows the elemental abundance ratios $^{36}\text{Ar}/^{132}\text{Xe}$ and $^{84}\text{Kr}/^{132}\text{Xe}$ for both the SNC materials and other meteorite groups for which reliable data exist (chondrites, ureilites, silicate inclusions in iron meteorites). In particular, note the positions in this plot of the C3V carbonaceous chondrites which, because of their high $^{129}\text{Xe}/^{132}\text{Xe}$ ratio^{21,22}, are the strongest and almost only candidates for martian atmosphere noble-gas carriers²⁰. Their position in Fig. 3 indicates that the fit is far from perfect. In fact, no known class of primitive meteorites seems to contain noble gases in the right proportions to qualify as an immediate source of the martian gases or those in SNC meteorites. Rather, there seems to be a dichotomy in the sense that chondrites and ureilites scatter around a straight line passing through the solar point while SNC meteorites, Mars and Earth form an indepen-

dent linear array with distinctly different slope and which definitely does not pass anywhere near the solar point. The generic relation between the noble gases in the atmospheres and rocks of evolved bodies suggests that the deviation from the solar pattern is a primary feature²³ related to either the building material or the accretion process. The similarity between the abundance patterns of Earth, Mars and SNC meteorites, in particular, makes it seem unlikely that adsorption of Xe on shales or ice should have played an essential part in converting the solar pattern into the planetary pattern on Earth²⁴.

While not impossible, a common parent body for all SNC meteorites is not easily reconciled with the data. If Nakhla and Chassigny are to be derived from the same parent body as the shergottite EETA 79001 (for which the evidence for a martian origin is strongest), processes have to be invoked which appear to be contrived. The martian thermal history in general and the evolution of its atmosphere in particular would have to have been profoundly different from what is believed to have been that of Earth.

We are greatly indebted to Drs Hutchison, Kurat and Pellas for making the meteorite specimens investigated available to us.

Received 26 February; accepted 8 July 1985.

1. Stolper, E., McSweeney, H. Y. Jr & Hays, J. F. *Geochim. cosmochim. Acta* **43**, 589–602 (1979).
2. Podosek, F. A. *Earth planet. Sci. Lett.* **19**, 135–144 (1973).
3. Papanastassiou, D. A. & Wasserburg, G. J. *Geophys. Res. Lett.* **1**, 23–26 (1974).
4. Gale, N. H., Arden, J. W. & Hutchison, R. *Earth planet. Sci. Lett.* **26**, 195–206 (1975).
5. Nakamura, N., Unruh, D. M., Tatsumoto, M. & Hutchison, R. *Meteoritics* **12**, 324–325 (1977).
6. Bogard, D. D. & Husain, L. *Geophys. Res. Lett.* **4**, 69–71 (1977).
7. Wooden, J. L. *et al. Lunar planet. Sci.* **10**, 1379–1381 (1979).
8. Bogard, D. D. & Nyquist, L. E. *Meteoritics* **14**, 356 (1979).
9. Shih, C.-Y. *et al. Geochim. cosmochim. Acta* **46**, 2323–2344 (1982).
10. Wood, C. A. & Ashwal, L. D. *Proc. lunar planet. Sci. Conf.* **12B**, 1359–1375 (1981).
11. Owen, T. *et al. J. geophys. Res.* **82**, 4635–4639 (1977).
12. Pollack, J. B. & Black, D. C. *Icarus* **51**, 169–198 (1982).
13. Bogard, D. D. & Johnson, P. *Science* **221**, 651–654 (1983).
14. Becker, R. H. & Pepin, R. O. *Earth planet. Sci. Lett.* **69**, 225–242 (1984).
15. Ganapathy, R. & Anders, E. *Geochim. cosmochim. Acta* **33**, 775–787 (1969).
16. Prior, G. *Mineralog. Mag.* **16**, 274–291 (1912).
17. Ozima, M. & Podosek, F. A. *Noble Gas Geochemistry* (Cambridge University Press, 1983).
18. Podosek, F. A. *Geochim. cosmochim. Acta* **34**, 341–365 (1970).
19. Jordan, J., Kirsten, T. & Richter, H. Z. *Naturforsch.* **35a**, 145–170 (1980).
20. Anders, E. & Owen, T. *Science* **198**, 453–465 (1977).
21. Matsuda, J.-I., Lewis, R. S., Takahashi, H. & Anders, E. *Geochim. cosmochim. Acta* **44**, 1861–1874 (1980).
22. Mazor, E., Heymann, D. & Anders, E. *Geochim. cosmochim. Acta* **34**, 781–824 (1970).
23. Podosek, F. A., Honda, M. & Ozima, M. *Geochim. cosmochim. Acta* **44**, 1875–1884 (1980).
24. Wacker, J. F. & Anders, E. *Geochim. cosmochim. Acta* **48**, 2373–2380 (1984).

Apparent icosahedral symmetry is due to directed multiple twinning of cubic crystals

Linus Pauling

Linus Pauling Institute of Science and Medicine,
440 Page Mill Road, Palo Alto, California 94306, USA

Recently the announcement was made of the remarkable discovery of intermetallic compounds with approximate composition MAl_6 ($\text{M} = \text{Cr}, \text{Mn}, \text{Fe}$) that formed crystals or pseudocrystals with icosahedral symmetry, as shown by the shape of the small nodules and by their electron diffraction patterns¹. I have found it hard to believe that any single crystal with 5-fold axes could give reasonably sharp diffraction patterns, resembling those given by crystals, and I have not been convinced to the contrary by the theoretical discussions of this possibility that have been published^{2–7}. I therefore set myself the task of predicting how a molten alloy of Mn (or Cr or Fe) and Al might react to sudden cooling. I have discovered that such an alloy on sudden cooling could form a metastable cubic crystal with a large cube edge, about 26.7 Å, with the unit cube containing about 1,120 atoms (possibly a few more), and that these crystals would show ordered multiple growth such that 20 of them, roughly tetrahedral in shape, grow out from a central seed in such a way as to produce an aggregate with approximately icosahedral symmetry.

The task of predicting a reasonable structure for this alloy was carried out with no information about the powder X-ray diffraction pattern except that one group of investigators had said that it could not be indexed by any Bravais lattice. The prediction of the structure was made entirely on the basis of knowledge of the effective radii of metal atoms and the principles determining the structure of metals and intermetallic compounds.

The transition metals form icosahedral complexes with Al; the crystal WAl_{12} , for example, is discussed in refs 8, 9. The effective radii⁸ (for ligancy 12) of Mn and Al are 1.268 and 1.429 Å, respectively, with ratio 0.89, corresponding closely to the value 0.902 for a regular icosahedron. Two MAl_{12} icosahedra might condense, with elimination of Al atoms, by sharing a corner, an edge or a face. Two icosahedra sharing an edge could easily eliminate another Al atom to share a face. I accordingly assumed that each MAl_{12} icosahedron would share four faces with surrounding icosahedra, leading to the composition MAl_6 . The four M—M vectors from each icosahedron to its neighbours are directed towards the corners of a regular tetrahedron. Moreover, when two icosahedra share a face, they are related by a plane of symmetry (the plane of the shared face). The other three faces of each icosahedron can be shared in either the eclipsed conformation or with a twist of 40°. The eclipsed conformation of tetrahedral bonds leads, with only slight deformation (reduction of the angle from 109.47°, the tetrahedral value, to 108°, the value for a pentagon), to pentagonal rings and then, when continued, to pentagonal dodecahedra of 20 icosahedra.

These structures are reminiscent of those of the rare-gas hydrates, which involve tetrahedrally hydrogen-bonded water molecules with the eclipsed relative orientation. The hydrate with a unit cube 12 Å on edge, containing 46 H_2O molecules, has 2 pentagonal dodecahedra and 6 tetrakaidecahedra in the unit¹⁰, whereas the one with unit cube 17 Å on edge, containing 136 H_2O molecules, has 16 pentagonal dodecahedra and 8 hexakaidecahedra in the unit¹¹. I selected the latter because of its larger number of dodecahedra.

Replacement of each water molecule by an MnAl_{12} icosahedron with shared faces leads to an infinite framework with 136 Mn and 816 Al atoms in the unit cube. This framework is similar to the framework of covalently bonded carbon atoms in a diamond crystal, with one body diagonal of each pentagonal dodecahedron in place of each C—C covalent bond.

I was able to make a moderately reliable prediction of the edge of the unit cube for the proposed MAl_6 structure. I assumed the Al—Al distance along the edges of the MAl_{12} icosahedron to be 2.864 Å, as in metallic Al. The distance between centres of icosahedra sharing a face is then 4.329 Å, 1.568 times the hydrogen-bond length in the 17-Å hydrate, with $a_0 = 17.0$ Å. With this factor the predicted value of the cube edge a of the alloy is 26.7 Å.

I obtained from D. Shechtman a copy of an X-ray diffractometric record, made with $\text{CuK}\alpha$ radiation in the US Bureau of Standards. This pattern shows 7 Si peaks (several with $\alpha_1\alpha_2$ splitting), 6 Al peaks (somewhat broader, indicating small crystallite size) and 22 other, rather broad, peaks, corresponding to the compound MnAl_6 . I found that all 22 peaks could be indexed with a cubic unit of structure. Moreover, the indicated value of the edge of the cubic unit of structure, 26.73 Å, turned out to be in good agreement with my predicted value, 26.7 Å. All of the observed diffraction maxima are those expected for spherically symmetrical groups about the diamond points 000, $\frac{1}{4}\frac{1}{4}\frac{1}{4}$, and so on.

The fact that the structure that I formulated for this alloy from chemical arguments with no knowledge about the X-ray diffraction diagram was found later to agree quantitatively with that diagram has convinced me that the structure is correct. I emphasize that I formulated only one structure, and that the X-ray diagram was not involved in any way in its formulation. The probability of chance agreement of the lattice constant to 0.1 Å and of adherence to the selection rules for intensities is surely less than 1 in 1,000.

Table 1 Analysis of X-ray diffraction pattern of icosahedrally twinned cubic crystals of MnAl_6

hkl	d (Å)	a (Å)	Intensity*
444	3.850	26.68	7
800	3.341	26.73	2
733	3.299	27.00	1
775	2.405	26.67	1
973, 1133	2.269	26.75	3
777, 1151	2.209	26.78	4
1222	2.168	26.73	100
1240	2.116	26.77	3
1082	2.060	26.70	150†
1511	1.7728	26.71	0.5
1460	1.7554	26.74	1
12124	1.5314	26.70	0.2
1680	1.4946	26.74	7
1684	1.4604	26.77	2
2040	1.3075	26.67	1
15119	1.2930	26.72	1
2062	1.2765	26.78	21
2440	1.0994	26.75	6
2444	1.0856	26.77	5
201212	1.0208	26.78	1
23117, 2575	1.0108	26.72	1
2488	1.0077	26.74	1
Average 26.73			

* All intensity values from 5 to 150 are those assigned by Shechtman and Blech². For most peaks the values of d are calculated from the values of 2θ printed on the diffractometer record. The exceptions are the weak reflections 4, 10, 11, 12 and 16.

† This very strong peak is about twice as broad as the peak with $d = 2.168$ Å. It is probably broadened by unresolved adjacent reflections 991 and 993.

The approximate intensity rules are given by the assumption of spherically symmetrical aggregates of atoms about the positions of the carbon atoms in diamond. They are that the reflections with h, k and l all even and their sum divisible by 4 should be the strongest, those with h, k and l all odd should be next, and those with h, k and l all even with sum not divisible by 4 the weakest. Table 1 verifies this prediction; the seven strongest reflections are of the first type, and none of the third type appears.

The absence or very low intensity of 111, 311, 331, 333 and 511 indicates that clusters of atoms, perhaps 13–20, occupy the centres of the hexakaidecahedra, around $\frac{1}{2}\frac{1}{2}\frac{1}{2}$ and so on. There are also smaller clusters, probably of four atoms, inside the pentagonal dodecahedra, increasing the number of atoms in the 25.73-Å cube to between 1,120 and 1,176. The composition of the alloys may be changed from MnAl_6 by these additional atoms.

An X-ray powder photograph taken with a focusing camera and X rays monochromatized by crystal reflection would probably show more than 100 powder lines, providing a more rigorous test of the proposed structure. Accurate values of predicted intensities of the lines will become available only after coordinates have been assigned to all of the atoms, a formidable task with such a complicated structure. This task is now being attempted (S. Samson, personal communication).

A further complication is that an icosahedron sharing four faces loses its symmetry operations of the second sort and becomes chiral, D or L. The eclipsed configuration described above requires that D and L alternate. Hence, in a ring of five icosahedra one DD or LL bond must occur, with a strain in the ring of 40°, which is a rotation of 4° per shared face in each icosahedron. The space-group symmetry is thus changed from that of diamond to a subgroup.

Although I knew about the work of Samson in determining the structures of NaCd_2 , Cu_4Cd_3 and $\beta\text{-Mg}_2\text{Al}_3$, I did not recognize until after my MnAl_6 structure had been formulated that they have essentially the same structure. The delay resulted

in part from the fact that in his descriptions of the structures Samson emphasized the Friauf polyhedra (ligancy 16 of the central large atom) rather than the icosahedra^{12–15}. I made an X-ray diffraction study of NaCd_2 in 1922 (ref. 16), but the structure was so complicated that I could not determine it then or during the next four decades, after which Samson was successful¹³. The edge of the unit cube is 30.56 Å, with between 1,120 and 1,190 atoms, not all of which were located by Samson. The structure contains 568 icosahedra, 124 Friauf polyhedra and 428–495 unidentified or others^{12,13}. A complete structure determination was made by Samson for Cu_4Cd_3 : edge of unit cube 25.87 Å, 1,124 atoms, 568 icosahedra, 124 Friauf polyhedra¹⁴, and a partial determination for $\beta\text{-Mg}_2\text{Al}_3$, edge of unit cube 28.24 Å, 1,192 atoms, 624–672 icosahedra¹⁵. The lattice constants differ from the value for MnAl_6 by amounts determined by the relative sizes of the atoms.

The first complex intermetallic compound found to have large clusters of atoms with local icosahedral symmetry was $\text{Mg}_{32}\text{Al}_{49}$, which has 162 atoms in a body-centred cubic unit¹⁷. The unit cube contains 98 icosahedra, 20 Friauf polyhedra and 44 others.

This compound gives an indication of the extent to which structures with local icosahedral symmetry can develop. It contains icosahedral groups of 117 atoms, a central atom with surrounding shells of 12, 32 and 72 atoms, the last being shared with surrounding groups (ref. 8, p. 428). The developing strain restricts the icosahedral complex to three shells of atoms around the central one. The hypothesis that I propose for the growth of a 20-fold twin of the MnAl alloy is that the seed is a pentagonal dodecahedron. It is surrounded by a shell of 12 dodecahedra, sharing faces, then by 32 (12 sharing faces with the inner 12, 20 sharing corners and with a body diagonal directed radially), and then by 72 (12 sharing faces with the next inner 12, 60 combining tetrahedrally with the next inner 20). At this point, the strain of attempting to fill space with pentagonal dodecahedra becomes great enough to disrupt the 12 outer dodecahedra, introducing additional atoms. The 20 tetrahedral groups of 4 dodecahedra continue to add dodecahedra in the way found in the 17-Å hydrate, and the twin of 20 pyramid-shaped cubic crystals with cube edge 26.73 Å develops.

The individual cubic crystals sometimes grow at equal rates, filling the spaces between them. Each assumes the shape of a trigonal pyramid, with the outer edges 5% longer than the radial edges. The multiple twin then has the macroscopic appearance of an icosahedron.

This postulated mechanism fixes the orientation of the 20 cubic crystals. Each crystal has a body diagonal extending out from a 3-fold axis of the seed dodecahedron and the other 3-fold axes lie in the planes of symmetry of the seed dodecahedron. A diffraction photograph of the twin would be the superposition of the diffraction photographs of the individual crystals that intercept the beam. I have not been able to make a thorough analysis of the published electron diffraction patterns because of lack of information about the conditions in which they were made, but my partial analysis indicates that the patterns are compatible with the proposed structure. The proposed structure also accounts for the dimensions of the pentagonal rings observed in electron micrographs of the alloys.

I conclude that this successful analysis of the diffraction pattern settles the question of the nature of the 'icosahedral' MnAl_6 alloys. They are multiple twins of a cubic crystal. The validity of this conclusion is strengthened by the fact that the value of the cube edge given by the diffraction pattern is almost exactly equal to that which I predicted, without any knowledge of this pattern, on the basis of purely chemical arguments. Crystallographers can now cease to worry that the validity of one of the accepted bases of their science has been questioned.

I thank Dr D. Schechtman for cooperation in providing me with the X-ray diffraction pattern and for other information, and Professor Barclay Kamb for pointing out to me that an icosahedron becomes chiral when it shares its faces. This investigation was supported in part by a grant from the Japan Shipbuilding Industry Foundation.

Received 21 June; accepted 4 September 1985.

1. Shechtman, D., Blech, I., Gratias, D. & Cahn, J. W. *Phys. Rev. Lett.* **53**, 1951 (1984).
2. Shechtman, D. & Blech, I. *Metal Trans.* (in the press).
3. Levine, D. & Steinhardt, P. J. *Phys. Rev. Lett.* **53**, 2477 (1984).
4. Bak, P. *Phys. Rev. Lett.* **54**, 1517 (1985).
5. Levine, D. *et al.* *Phys. Rev. Lett.* **54**, 1520 (1985).
6. Mermin, N. D. & Trorin, S. M. *Phys. Rev. Lett.* **54**, 1524 (1985).
7. Heiney, P. A. *Nature* **315**, 178 (1985).
8. Pauling, L. *The Nature of the Chemical Bond* 3rd edn (Cornell University Press, Ithaca, 1960).
9. Adam, J. & Rich, J. B. *Acta crystallogr.* **7**, 813 (1954).
10. Pauling, L. & Marsh, R. E. *Proc. natn. Acad. Sci. U.S.A.* **36**, 112 (1952).
11. Claussen, W. F. *J. Chem. Phys.* **19**, 259, 662, 1425 (1951).
12. Samson, S. in *Structural Chemistry and Molecular Biology* (eds Rich, A. & Davidson, N.) 687-717 (Freeman, San Francisco, 1968).
13. Samson, S. *Nature* **195**, 259-262 (1962).
14. Samson, S. *Acta crystallogr.* **17**, 491 (1964).
15. Samson, S. *Acta crystallogr.* **19**, 401 (1965).
16. Pauling, L. *J. Am. chem. Soc.* **45**, 2777 (1923).
17. Bergman, G., Waugh, J. L. T. & Pauling, L. *Nature* **169**, 1057-1058 (1952); *Acta crystallogr.* **10**, 254 (1957).

Momentum flux in breaking waves

W. K. Melville & Ronald J. Rapp

Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139, USA

Wave breaking is believed to be important in air-sea interaction. Laboratory measurements¹ suggest that the momentum flux from the atmosphere to the ocean may be significantly enhanced by breaking and recent field measurements²⁻⁴ have demonstrated the important role of breaking in bubble generation and gas transfer. It has long been speculated that the loss of momentum flux from the wave field due to breaking could act as a source of momentum for current generation⁵ and Mitsuyasu⁶ has drawn attention to the large discrepancy between the momentum flux from the wind and that carried by the waves, suggesting that the loss may be due to wave breaking. Here we present what we believe are the first well-controlled laboratory measurements of the momentum flux lost by wave breaking. These measurements are consistent with Mitsuyasu's hypothesis and recent measurements of wave growth, and support the conclusion that wave breaking plays an important role in momentum transfer across the air-sea interface.

Wave breaking may result from intrinsic instabilities of uniform surface waves^{7,8}, wave-current interaction⁹, and the superposition and nonlinear interaction of spectral components of a random wave field. The breaking and white-capping associated with the modulation of the waves near the peak of the spectrum^{3,10}, (that is, with the tendency of the waves to occur in groups) is likely to be due to the latter mechanism. We simulated this condition in a wave channel^{5,11} in which single isolated packets of waves were generated by superposition of Fourier components of equal amplitude a_0 at the wavemaker in a radian frequency band $(f_0 \pm \Delta f/2)$. The initial phase for each component was chosen to give constructive interference of linear surface waves at a point x_0 down the channel. This gives a nominal dimensionless duration of the packet at breaking of $2f_0/\Delta f$ wave periods. Dimensional reasoning then implies that the dimensionless wave amplitude a/a_0 and other variables will be functions of $(xk_0, t f_0)$ with parametric dependence on $(a_0 k_0, \Delta f/f_0, x_0 k_0)$ where x is the distance down the channel, t is time and $k_0 = f_0/g$, the linear dispersion relationship for deep-water waves.

The parameters $a_0 k_0$, $\Delta f/f_0$, $x_0 k_0$, which are dimensionless measures of amplitude, bandwidth and phase, were varied in the ranges 0.1-0.5, 0.4-1.4 and 27.4-66.0, respectively. The water depth was 60 cm and f_0 was in the range 5.5-8.1 rad s⁻¹.

Figure 1 shows an example of the evolution of a wave packet at various stations down the channel. The incidence of breaking depends on the parameters above, most notably a_0 . Figure 2 shows examples of single spilling and plunging events at specific values of $a_0 k_0$. At other values of $a_0 k_0$, multiple breaking waves were observed, and this led to consideration of losses due to

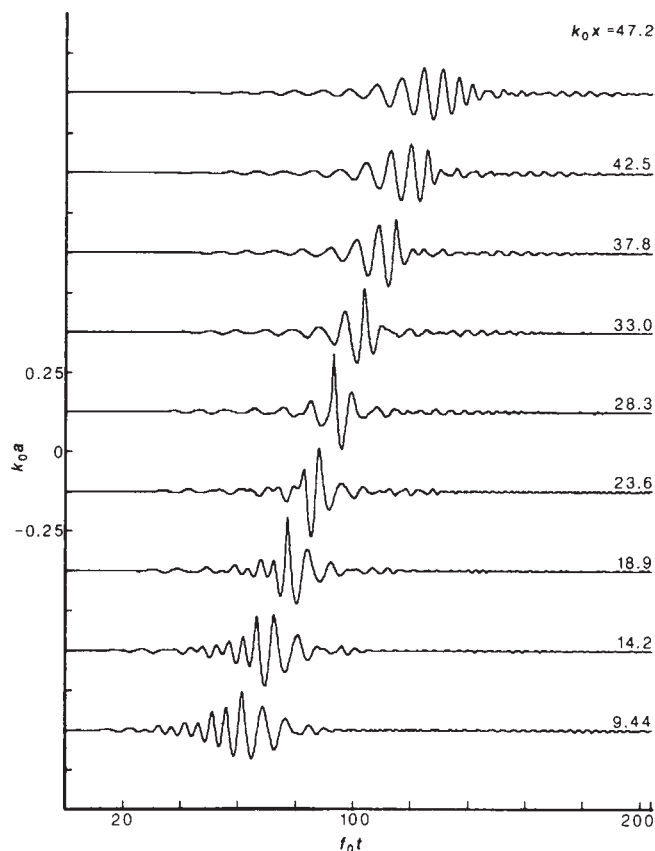


Fig. 1 Time series of surface displacement at stations down the wave channel for a spilling breaker: $f_0 = 6.79 \text{ s}^{-1}$, $\Delta f/f_0 = 0.73$, $x_0 k_0 = 27.4$, $a_0 k_0 = 0.30$. The time is normalized by f_0 and the amplitude by k_0 . Breaking occurs at $xk_0 = 28$, where xk_0 is the normalized distance from the paddle.

'breaking' of the packet rather than single breaking waves. Indeed, multiple breaking events are observed in the field³. In these experiments the breaking process (whether single or multiple) was taken to be a 'black box' and the loss of excess momentum flux was evaluated by measuring the flux upstream and downstream of the breaking region.

For weakly nonlinear, slowly varying two-dimensional deep-water waves the excess momentum flux is proportional to a^2 (ref. 12). Thus we integrated a^2 during passage of the packet at various stations down the channel. Examples of this measure of integrated momentum flux, S , with distance down the channel are shown in Fig. 3 for incipient breaking, spilling and plunging events. The loss of momentum flux, ΔS , due to breaking is determined by taking the difference between the incipient breaking and breaking measurements downstream. The breaking region is clearly shown in this figure by a rapid decrease in S/S_0 , the dimensionless excess momentum flux. The decline in S/S_0 prior to breaking is, we believe, due to the increasing influence of nonlinearity and strong modulation of the wave variables as the breakpoint is approached. Our use of the downstream differences, where S/S_0 is essentially constant, avoids its explicit consideration. We believe the oscillations in the breaking region are caused by rapid modulations of the wave variables which, as a consequence of dispersion, do not occur far upstream and downstream of this region. They are not accounted for in our approximation of the momentum flux, and are not relevant for our present purposes.

Results like those of Fig. 3 were collected for a range of parameters and used to determine the excess momentum flux loss due to breaking, $\Delta S/S_0$, as a function of $a_0 k_0$, $\Delta f/f_0$, $x_0 k_0$. An example is shown in Fig. 4 for fixed $x_0 k_0$ for different amplitudes and bandwidths. This figure is typical of the data

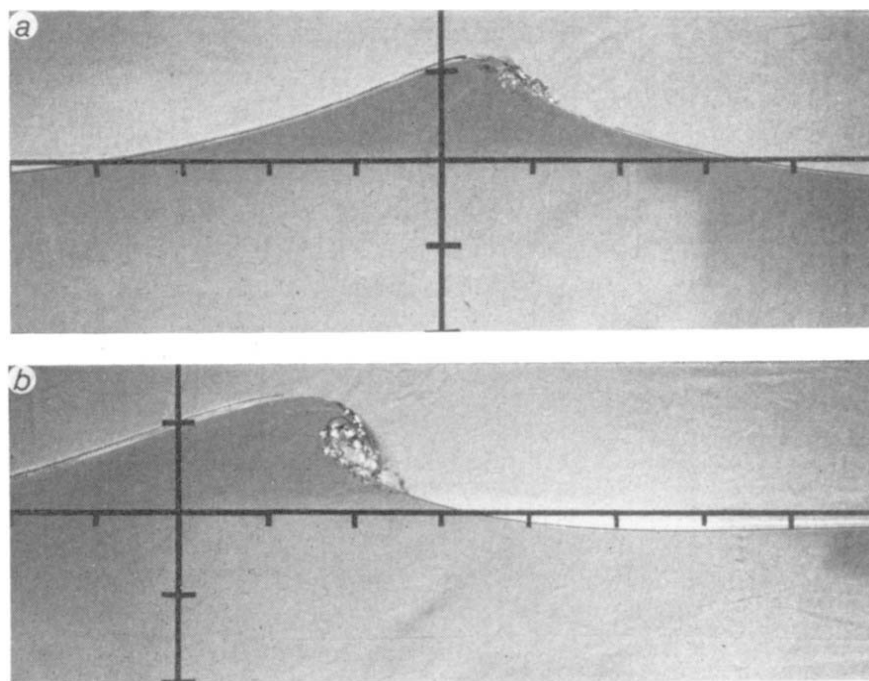


Fig. 2 Examples of single breaking waves for $f_0 = 5.53 \text{ s}^{-1}$, $\Delta f/f_0 = 0.73$, $x_0 k_0 = 27.4$. *a*, Spilling, $a_0 k_0 = 0.28$. *b*, Plunging, $a_0 k_0 = 0.35$. The marks are spaced at 10-cm intervals, and the upper boundary of the horizontal bar is at the still-water level.

Fig. 3 Variance of the free surface displacement normalized by its upstream reference value at stations down the channel: $f_0 = 6.79 \text{ s}^{-1}$, $\Delta f/f_0 = 0.73$, $x_0 k_0 = 27.4$. The amplitude parameter $a_0 k_0$ is: 0.26, incipient breaking, +; 0.30, single spill, Δ ; 0.39, single plunger, $\square$. The normalized variance is approximately proportional to the excess momentum carried by the wave field past the measuring station, upstream and downstream of the breaking region. The actual breaking locations are marked by S (spilling) and P (plunging). The run denoted by Δ is that shown in Fig. 1. The scatter in repeated measurements was within the size of the symbols.

Table 1 Estimates of $\Delta S/S$, the fractional increase of wave momentum flux due to wind forcing between breaking events

U_{10} (m s^{-1})	u_* (m s^{-1})	T (s)	c (m s^{-1})	u_*/c	$\bar{x}/\lambda$	$\Delta S/S$
6	0.21	2.6	4.1	0.05	20	0.04
12	0.48	3.1	4.8	0.10	13	0.09

Values determined according to equation (3) and data of refs 3 and 14. Variables are defined in the text.

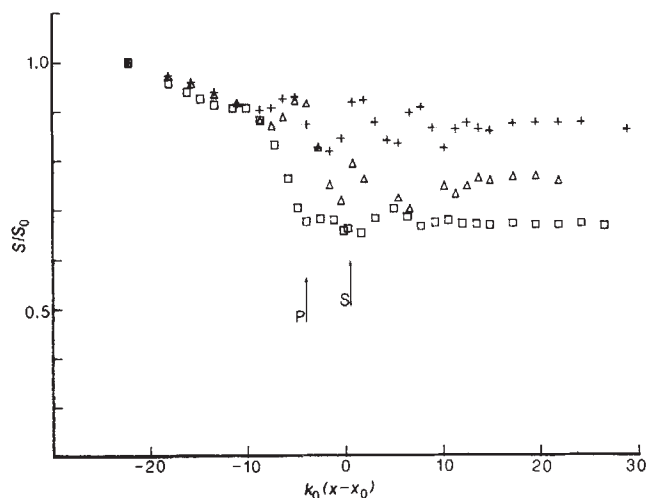
collected and shows that the losses increase rapidly with increasing amplitude: 10% or more of the momentum flux carried by the packet may be lost through one breaking wave, and up to 20–30% from one multiple breaking event. The experiments show that the fractional loss of excess momentum from one wave packet due to one breaking event is in the range

$$\frac{\Delta S}{S_0} = 0(10^{-2} - 10^{-1}) \quad (1)$$

From wave growth measurements in deep water, Mitsuyasu⁶ found that the divergence of the wave momentum flux inferred from direct measurements of wave growth in the laboratory is given by

$$\frac{dS}{dx} = 0.68 \left[\frac{u_*^2}{c} \right] \frac{S}{\lambda} \quad (2)$$

where u_* is the friction velocity of the wind, c the phase speed of the waves, λ is wavelength and $0.1 \leq u_*/c \leq 1.0$. The coefficient 0.68 is within the range (0.5 ± 0.25) inferred by Plant's¹³ extensive review of laboratory and field data. Mitsuyasu



went on to show that dS/dx derived from the fetch relation for wind waves is an order of magnitude less than that given by equation (2), and suggested that a large fraction of this equation must then be balanced by wave breaking.

Now if $\Delta S/S \ll 1$, we may approximate equation (2) by its difference form

$$\frac{\Delta S}{S} = 0.68 \left[\frac{u_*^2}{c} \right] \frac{\Delta x}{\lambda} \quad (3)$$

If $\Delta S/S$ over the distance between breaking events is comparable in magnitude to the fractional loss due to a breaking event, then we may state that the momentum lost due to breaking is comparable to that gained from the wind. As far as we are aware, the only measurements of the distance between breaking events (in the direction of the wind) are those of Thorpe and Humphries³ who plotted $\lambda/\bar{x}$, the ratio of wavelength to mean distance between breaking waves, or the number of breaking waves per wave, versus the windspeed U_{10} , for $5 < U_{10} < 15 \text{ m s}^{-1}$. In this speed range, they found that $0.02 < \lambda/\bar{x} < 0.19$. Unfortunately, they only gave the phase speed c at two values of U_{10} and no explicit information on u_* . We may proceed to

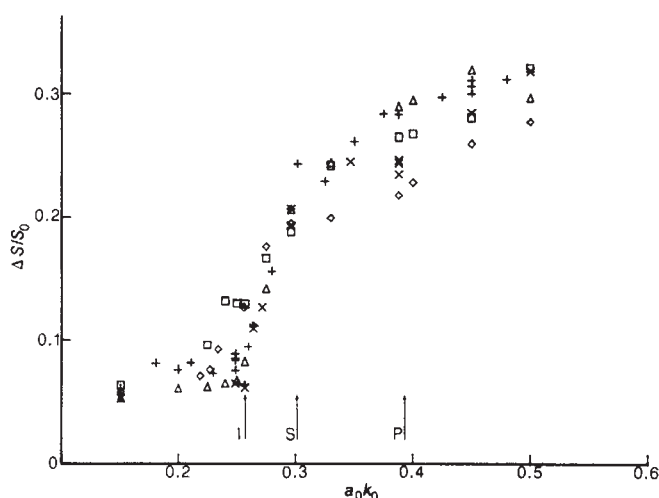


Fig. 4 Excess momentum lost from the wave field due to breaking as a function of the amplitude parameter, a_0k_0 , for $\Delta f/f_0 = 0.4$ ($\square$), 0.6 ($\triangle$), 0.73 (+), 1.0 ($\times$), 1.4 ($\diamond$); $f_0 = 6.79 \text{ s}^{-1}$, $x_0k_0 = 27.4$. The arrows show the amplitudes for incipient breaking (I), single spill (S) and single plunger (P) for $\Delta f/f_0 = 0.73$.

use their data for these two points by making one assumption: that the atmospheric conditions were near neutral. We may then use Wu's¹⁴ correlation of drag coefficients to estimate u_* from U_{10} . This we have done, and the results show that according to equation (3), $\Delta S/S$ between breaking events is in the range 0.04–0.09 for windspeeds in the range 6–12 m s^{-1} (Table 1). Thorpe and Humphries' data further imply that $\bar{x}$, their mean distance between breaking events, is comparable to the group length (see Fig. 3d of ref. 3).

Our laboratory measurements (Fig. 4) show that $\Delta S/S$ is approximately 0.1 for a single spilling break, where S here refers to the excess momentum flux carried by one group. The important feature of Fig. 4 is the rapid initial increase in $\Delta S/S$ with ak once the breaking threshold is passed. This implies that even the gentle breaking events contribute significant fractional losses in the momentum flux, comparable to those for a single spilling break.

We conclude that the excess momentum flux lost due to wave breaking may be comparable to that transferred from the wind. Further, our measurements imply that in future quantitative studies of wave breaking, it may be most fruitful to consider the dynamics of the breaking wave group rather than considering single breaking waves.

We thank Mr Eng-Soon Chan for collaboration on preliminary measurements and Dr Martin Greenhow for advice on the generation of breaking waves. This work is supported by NSF grants OCE-8214746 and MEA-8210169.

Received 28 May; accepted 25 July 1985.

- Banner, M. L. & Melville, W. K. *J. Fluid Mech.* **77**, 825–842 (1976).
- Thorpe, S. A. & Stubbs, A. R. *Nature* **279**, 403–405 (1979).
- Thorpe, S. A. & Humphries, P. N. *Nature* **283**, 463–465 (1980).
- Thorpe, S. A. *Phil. Trans. R. Soc. A* **304**, 155–210 (1982).
- Longuet-Higgins, M. S. in *10th Symp. Naval Hydrodynamics*, 597–605, ONR, Arlington, Virginia (US Government Printing Office, 1976); in *Turbulent Flusses through the Sea Surface, Wave Dynamics, and Prediction*, (eds Favre, A. & Hasselmann, K.) 199–220 (Plenum, New York, 1978).
- Mitsuyasu, H. *J. geophys. Res.* **90**, 3343–3345 (1985).
- Longuet-Higgins, M. S. & Cokelet, E. D. *Proc. R. Soc. A* **350**, 1–25 (1976).
- Melville, W. K. *J. Fluid Mech.* **115**, 165–185 (1982).
- Banner, M. L. & Phillips, O. M. *J. Fluid Mech.* **65**, 647–656 (1974).
- Donelan, M., Longuet-Higgins, M. S. & Turner, J. S. *Nature* **239**, 449–451 (1972).
- Melville, W. K., Rapp, R. J. & Chan, E. S. in *The Ocean Surface: Wave Breaking, Turbulent Mixing and Radio Probing* 413–418 (Reidel, Dordrecht, 1985).
- Whitham, G. B. *Linear and Nonlinear Waves* (Wiley, New York, 1974).
- Plant, W. J. *J. Geophys. Res.* **87**, 1661–1667 (1982).
- Wu, J. *J. geophys. Res.* **87**, 9704–9706 (1982).

Magnetic differentiation of atmospheric dusts

F. Oldfield*, A. Hunt*, M. D. H. Jones*, R. Chester†, J. A. Dearing‡, L. Olsson§ & J. M. Prospero||

Departments of *Geography and †Oceanography, University of Liverpool, PO Box 147, Liverpool L68 3BX, UK

‡ Department of Geography, Coventry Polytechnic, Coventry CV1 5FB, UK

§ Department of Physical Geography, University of Lund, Sölvegatan 13, S-22362 Lund, Sweden

|| Department of Marine and Atmospheric Chemistry, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149, USA

Although several magnetic separation methods have been used to provide some initial subdivision and characterization of the particulate emission products of fossil-fuel combustion^{1–5}, only recently has there been any attempt to use magnetic properties to distinguish between different emission types^{6,7} and atmospheric particulate sources⁸. In the present study, we use several concentration-independent magnetic parameters and ratios to characterize dusts from different source types. Non-destructive measurements of saturation isothermal remanence, anhysteretic remanence and frequency-dependent susceptibility differentiate atmospheric dusts into groups characterized by distinctive magnetic mineral and grain size assemblages. These assemblages can be related to the source of the dusts as they reflect differences in the conditions under which the magnetic oxides developed. The magnetic parameters are sensitive to differences between dusts arising from fossil-fuel combustion and from other industrial processes, and those derived from soil erosion. Within the set of soil-derived dusts, the magnetic parameters distinguish between the weathering regimes operating in different source areas.

Chester *et al.*⁸ were able to distinguish between and characterize dust source types in the Mediterranean region by plotting aluminium concentration against the magnetic susceptibility/aluminium concentration quotient. Soil-derived dusts of North African origin fell at one extreme, with high aluminium concentrations and a low quotient, whereas urban/industrial dusts rich in magnetic spherules^{9–11} fell at the opposite extreme. On a local scale, Hunt *et al.*⁷, using magnetic parameters alone, were able to distinguish between fly ash originating from power stations and the particulates in automobile emissions. They were able to show consistent differences related to the coercivity of saturation isothermal remanence (see Fig. 1 legend). Such parameters, together with a range of interparametric ratios, have been used successfully to identify shifts in sediment source types in rivers^{12,13}, lakes¹⁴, and a range of near-shore marine environments^{15–17}; they are independent of concentration and can usually be measured rapidly and non-destructively on samples of 0.05 g or less.

Table 1 summarizes some of the measurements obtained for samples collected between October 1966 and July 1969 from the eastern tip of Barbados (13°10' N, 59°30' W) by means of 1 m × 0.5 m nylon mesh panels mounted on a 14-m-high platform and arranged so as to face the prevailing wind at all times^{18,19}. Delaney *et al.*¹⁸ regard the sampling efficiency of the method as being low for particles <1 μm in diameter. Out of a total of 20 samples received for magnetic measurement, four 'between-season' samples from October and November were omitted. The rest were identified as either red-brown, Sahara- and Sahel-derived 'summer' dusts, or grey, more locally derived, South American 'winter and spring' dusts^{19,20}. The two populations are clearly distinguished on the basis of the parameters tabulated. For the Mann-Whitney U statistic, they show a significant difference at the $P = 0.05$ rejection level. The high coercivities of remanence (B_{CR}), 'harder' isothermal remanent magnetization (IRM)/saturation IRM (SIRM) quotients and higher SIRM/ χ (low-field susceptibility) values for the Saharan set are consistent with a high haematite contribution typical of surface

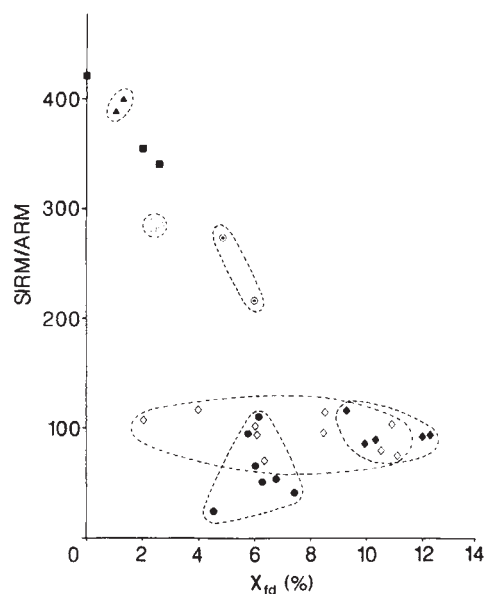


Fig. 1 SIRM/ARM versus χ_{fd} % for various dust sets from Barbados, North Atlantic and the inland sea of Japan. Resuspended particle-sized fly ash (■) is included for comparison. ▲, Inland Sea of Japan dusts ($SIRM > 50 \times 10^{-3} \text{ A m}^2 \text{ kg}^{-1}$); ○, North Sea and North Atlantic dusts ($SIRM > 50 \times 10^{-3} \text{ A m}^2 \text{ kg}^{-1}$, that is, most 'polluted'); ●, North Atlantic dusts ($SIRM = 21 \times 10^{-3} \text{ A m}^2 \text{ kg}^{-1}$); ◆, Barbados 'summer' dusts ($SIRM = 4-8 \times 10^{-3} \text{ A m}^2 \text{ kg}^{-1}$); ◇, Barbados 'winter' dusts ($SIRM = 4-8 \times 10^{-3} \text{ A m}^2 \text{ kg}^{-1}$).

Methods. Saturation isothermal remanent magnetization (SIRM) was grown in the present samples by means of a Molspin pulse magnetizer generating a maximum d.c. field of 0.82 T. Reverse-field ratios ($IRM_{-20 \text{ mT}}/SIRM$, etc.—see Table 1) and the coercivity of SIRM (B_{CR}) were determined by placing previously 'saturated' and measured samples in a succession of increasing reverse fields generated by a less powerful but more precisely adjustable pulse magnetizer. More strongly positive reverse-field ratios and higher B_{CR} values indicate that the initial forward remanence is more difficult to reverse. Failure of the isothermal remanence to reverse saturate in fields in excess of 200 mT points to a significant anti-ferrimagnetic (haematite) contribution to the initial forward remanence. Anhysteretic remanent magnetization (ARM) was grown by subjecting the sample to a d.c. field of 0.05 mT superimposed on an a.c. field smoothly declining from a peak of 100 mT generated in an a.c. demagnetizing unit. Stable single-domain ferrimagnetic grains are characterized by relatively high ARM values^{26,28}. All remanences were measured using a portable Mini-spin fluxgate magnetometer with a noise level of $\sim 1 \times 10^{-6} \text{ A m}^2$. Low-field susceptibility (χ) was measured using the low-frequency (0.5 kHz) circuit of a Bartington susceptibility meter with a noise level of $\sim 1 \times 10^{-8} \text{ SI units}$. χ is usually considered to be proportional to the concentration of ferrimagnetic minerals in a sample, although in soil-derived material especially, it is also often influenced by the presence of paramagnetic forms of iron. Frequency-dependent susceptibility (χ_{fd}) was measured using a Bartington dual-frequency sensor. χ_{fd} is the difference between the low-frequency (0.5) and high-frequency (5 kHz) measurements, expressed as a percentage of the low-frequency measurements. In the case of the present samples, the denominator (χ) must be adjusted for diamagnetic effects from the sample holder and packing. The frequency-dependent component of total low-field susceptibility reflects the presence of ferrimagnetic grains with a diameter just below the lower limit for stable single-domain grains ($\sim 0.03 \mu\text{m}$)³¹.

materials in arid and semi-arid regions²¹. In the magnetically 'hardest' samples, up to 30% of the original SIRM remains unsaturated in a reverse field of 400 mT. The grey dusts have, on average, a higher frequency-dependent component (χ_{fd} %) in their total low-field susceptibility (Fig. 1) and are almost entirely reverse-saturated at fields $< 400 \text{ mT}$. All the values are

consistent with a derivation largely from fine secondary ferrimagnetic grains at the surface of magnetically 'enhanced'²², possibly burnt^{23,24} soils (see below).

Figure 1 plots SIRM/anhysteretic remanent magnetization (ARM) versus frequency-dependent susceptibility as a percentage of total low-field susceptibility (χ_{fd} %) for a larger sample set including not only the Barbados ones but also several samples from North Sea and North Atlantic cruises as well as two samples from the inland sea of Japan. The Barbados dusts are grouped into seasonal sets as in Table 1. The cruise samples were collected on Terylene meshes mounted onboard research vessels during 1980 and 1981. Each sample reflects a collection period of several days, using two 1-m^2 meshes mounted 4 m above the prow of the vessel. Chester *et al.*²⁵ describe the method fully and suggest that the particles retained are predominantly greater than $5 \mu\text{m}$.

The samples from the Atlantic can be placed in three well-defined groups according to their SIRM values. The gradient from highest to lowest SIRM reflects the declining relative importance of anthropogenic sources⁸ within the samples, which span a large area from the coasts of Britain to low latitudes comparable to those of Barbados. The most heavily polluted samples from the North Sea and Bay of Biscay are largely of

Table 1 Selected magnetic parameters for the Barbados dusts

	$SIRM/\chi$ (kA m^{-1})	$IRM_{LR}/SIRM$	$IRM_{HR}/SIRM$	B_{CR} (mT)
Grey dusts (December–April)	8.0 ± 0.23	0.24 ± 0.06	-1.44 ± 0.09	35 ± 1.7
Red-brown dusts (May–September)	9.25 ± 0.63	0.45 ± 0.11	-1.22 ± 0.13	42 ± 4.1

(Values are mean ± 1 s.d.)

$$IRM_{LR}/SIRM = \frac{IRM_{-20 \text{ mT}}}{SIRM} + \frac{IRM_{-40 \text{ mT}}}{SIRM}$$

$$IRM_{HR}/SIRM = \frac{IRM_{-100 \text{ mT}}}{SIRM} + \frac{IRM_{-200 \text{ mT}}}{SIRM}$$

The parameters shown are all independent of concentration. The $IRM/SIRM$ ratios have been chosen to summarize the demagnetization characteristics of the samples in relatively low (LR) and high (HR) reverse fields (see Fig. 1 legend). The red-brown dusts have a higher $SIRM/\chi$ quotient and a significantly 'harder' isothermal remanence.

north-west European origin; they have maximum SIRM values and also have the highest SIRM/ARM ratio and the lowest χ_{fd} % values. The least heavily polluted samples, with minimum SIRM, have the lowest SIRM/ARM ratios and consistently higher χ_{fd} % values.

The ratios in the 'dirty' set come close to the range of values for power station fly ash, whereas those for the 'cleanest' set overlap with the Barbados summer dusts. The latter were low-latitude north-east trade wind dusts of probable North African origin.²⁶ The ratios for the inland sea of Japan samples lie within the 'fly ash' range, and scanning electron microscope examination of magnetic extracts from these samples confirms that they are rich in spherules (A.H., unpublished data). The basis for the magnetic differentiation shown in Fig. 1 is variation in magnetic grain (that is, crystal) not particle, size. Banerjee *et al.*²⁷ and King *et al.*²⁸ have shown that ARM values are relatively high in grain size assemblages rich in true stable single-domain crystals ($\sim 0.05 \mu\text{m}$ diameter). King *et al.*²⁸ ascribe high ARM values to surface soil-derived material in which the stable single-domain crystals are characteristic of the secondary ferrimagnetic oxides resulting from magnetic enhancement²⁹; this is consistent with a wide range of evidence now available^{21,30}. Relatively low ARM values and high SIRM/ARM ratios are common to all the anthropogenic particulates studied so far (A.H., unpublished data). The ordinate of Fig. 1 thus portrays changes in SIRM/ARM expressing the control over magnetic grain size assemblage exercised by the shift in source type between predominantly fossil-fuel combustion-derived and soil

erosion-derived particulates. Variations along the abscissa are also related to grain size as the frequency-dependent component of total magnetic susceptibility is related to the presence of ferrimagnetic crystals at the stable single-domain, superparamagnetic boundary ($\sim 0.03 \mu\text{m}$ diameter)³¹.

These results, together with those published previously^{7,8}, suggest that magnetic parameters can be of considerable value in characterizing atmospheric dusts and aerosols derived from different sources, whether the differentiation results from seasonal variations at a single collection site or large-scale spatial variations. Statistical treatment using discrimination analysis applied to a wider range of measurements on a larger dust set, confirms the classificatory value of magnetic measurements³². Comparable results from Pacific Island dust sets (A.H., unpublished data), mesh samples obtained from most of the world's oceans³³ and from growing peat bogs, which retain a record of the history of particulate deposition^{34,35}, suggest that the present approach to characterizing variations in atmospheric particulate types will be applicable to a wide range of temporal and spatial scales.

Received 8 May; accepted 22 July 1985.

1. Theis, T. L. & Wirth, J. L. *Envir. Sci. Technol.* **11**, 1095-1100 (1977).
2. Hansen, L. D., Silberman, D. & Fisher, G. L. *Envir. Sci. Technol.* **15**, 1057-1062 (1981).
3. Linton, R. W., Natusch, D. F. S., Solomon, R. L. & Evans, C. A. Jr *Envir. Sci. Technol.* **14**, 159-164 (1980).
4. Olson, K. W. & Skogerboe, R. K. *Envir. Sci. Technol.* **9**, 227-230 (1975).
5. Hulett, L. D., Weinberger, A. J., Ferguson, M., Northcutt, K. J. & Lyon, W. S. DE-8102855 EPR-IA-1822 (1981).
6. Chaddha, G. & Seehra, M. S. *J. Phys. D* **16**, 1767-1776 (1983).
7. Hunt, A., Jones, J. M. & Oldfield, F. *Sci. Total Envir.* **33**, 129-139 (1984).

8. Chester, R., Sharples, E. J., Sanders, G., Oldfield, F. & Saydem, A. C. *Wat. Air Soil Pollut.* **23**, 25-35 (1984).
9. Doyle, J. L., Hopkins, T. L. & Betzer, P. R. *Science* **194**, 1157-1159 (1976).
10. Oldfield, F., Thompson, R. & Barber, K. E. *Science* **199**, 679-680 (1978).
11. Puffer, J. H., Russell, E. W. B. & Rampino, M. R. *J. sedim. Petrology* **50**, 247-256 (1980).
12. Oldfield, F., Rummery, T. A., Thompson, R. & Walling, D. E. *Wat. Resources Res.* **15**, 211-218 (1979).
13. Walling, D. E., Peart, M. R., Oldfield, F. & Thompson, R. *Nature* **281**, 110-113 (1979).
14. Oldfield, F., Maher, B., Higgitt, S. R. & Scoullios, M. *EEC Desertification Conf. Proc.*, Mytilini (1984).
15. Oldfield, F. in *Background to Palaeohydrology* (ed. Gregory, K. J.) 141-165 (Wiley, London, 1983).
16. Oldfield, F. & Scoullios, M. *Mar. Pollut. Bull.* **15**, 229-231 (1984).
17. Oldfield, F. & Maher, B. *Conf. Proc. Drainage Basin Erosion and Sedimentation* (ed. Loughran, R.) (Newcastle, New South Wales, Australia, 1984).
18. Delaney, A. C., Parkin, D. W., Griffin, J. J., Goldberg, E. D. & Reimann, B. F. *Geochim. cosmochim. Acta* **31**, 885-909 (1967).
19. Prospero, J. M. *Bull. Am. met. Soc.* **49**, 645-672 (1968).
20. Prospero, J. M., Glaccum, R. A. & Nees, R. T. *Nature* **289**, 570-577 (1981).
21. Maher, B. thesis, Univ. Liverpool (1984).
22. Mullins, C. F. *J. Soil. Sci.* **28**, 223-246 (1977).
23. Le Borgne, E. *Ann. Geophys.* **11**, 399-419 (1955).
24. Rummery, T. A., Bloemendal, J., Dearing, J. A., Oldfield, F. & Thompson, R. *Ann. Geophys.* **35**, 103-107 (1979).
25. Chester, R., Elderfield, H., Griffin, J., Johnson, L. & Padgham, R. *Mar. Geol.* **13**, 91-105 (1972).
26. Parkin, D. W., Phillips, D. R., Sullivan, R. A. L. & Johnson, L. R. *Q. Jl. R. met. Soc.* **98**, 798-808 (1972).
27. Banerjee, S. K., King, J. & Marvin, J. *Geophys. Res. Lett.* **8**, 333-336 (1981).
28. King, J., Banerjee, S. K., Marvin, J. & Ozdemir, O. *Earth planet. Sci. Lett.* **59**, 404-419 (1982).
29. Longworth, G. et al. *J. Soil. Sci.* **30**, 93-110 (1979).
30. Thompson, R. & Oldfield, F. *Environmental Magnetism* (George Allen & Unwin, London, 1985).
31. Mullins, C. E. & Tite, M. S. *J. geophys. Res.* **78**, 804-809 (1973).
32. Hunt, A. *Physics Earth planet. Interiors* (in the press).
33. Chester, R. in *The Changing Chemistry of the Oceans* (eds Dyssen, D. G. & Jagger, P.) 20th Nobel Symposium Volume, 1972).
34. Oldfield, F. et al. *Conf. Proc. Palaeolimnology and Acidification* (ed. Norton, S.) 409-441 (Bangor, Maine, 1984).
35. Oldfield, F., Tolonen, K. & Thompson, R. *Ambio* **10**, 185-188 (1981).

³He/⁴He ratios of marine ferromanganese nodules

Yuji Sano, Kazuhiro Toyoda & Hiroshi Wakita

Laboratory for Earthquake Chemistry, Faculty of Science, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

Helium isotope ratios in the Earth's materials provide useful information about their mechanism of formation. Although more than 1,000 measurements have been reported, most are for land areas, and little is known about helium ratios on the ocean floor. Here we present new He isotope data for ten marine ferromanganese nodules, together with isotope compositions for Ar and chemical compositions. The observed ³He/⁴He ratios are extraordinarily high, with values up to 59.3×10^{-6} . Considering that the direct supply of materials from the deep mantle is an insignificant factor in the area studied, the negligible contribution of radiogenic ³He, and the improbability of artificial nuclear fallout debris, the most likely explanation appears to be conglomeration of extraterrestrial matter. Abundances of radiogenic ⁴⁰Ar are related to the genetic environments of nodules (that is, diagenetic or hydrogenetic) as characterized by their chemical compositions.

Based on their physical and chemical signatures, marine ferromanganese nodules are generally considered to be generated in two major environments¹⁻³: diagenetic and hydrogenetic. The diagenetic type is characterized by a high Mn-to-Fe ratio (Mn/Fe ≥ 2.5), enrichment of trace elements such as Co, Ni and Cu, rough surface structure and the presence of 10-Å manganite in X-ray diffraction patterns. Nodules of this type are mainly attributed to upward diffusion of Mn from diagenetic pore water in sediments and to precipitation in an oxidized surface zone. The hydrogenetic type is characterized by a low Mn-to-Fe ratio (Mn/Fe ≤ 2.5), relative depletion of trace elements, smooth surface structure and the presence of δ -MnO₂. Nodules of this type grow by very slow accumulation of inorganic colloidal particles of hydrated metal oxides from near-

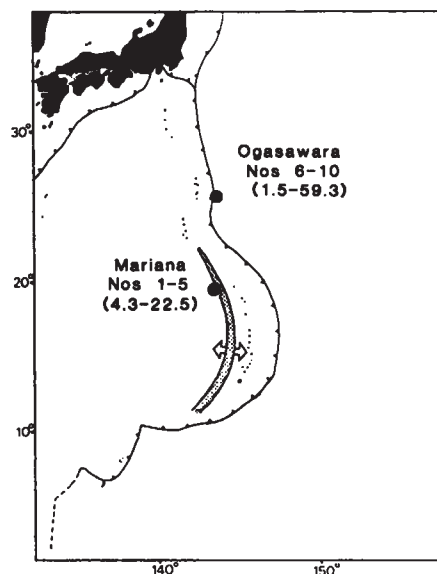


Fig. 1 Sampling sites of marine ferromanganese nodules. Figures in parentheses indicate ³He/⁴He ratios in units of 10^{-6} .

bottom seawater¹⁻³. Contribution from submarine volcanoes has also been suggested for this type⁴.

Isotope compositions of noble gases, particularly ³He/⁴He ratios, provide constraints on the genetic sources for Earth's materials^{5,6}. It has been confirmed that pore water in deep sea sediments contains radiogenic He (refs 7-9); nevertheless submarine volcanoes discharge mantle-derived He (ref. 10). If nodules are diagenetic, the contribution of the radiogenic He from pore water will be higher. If they are hydrogenetic, the atmospheric and/or mantle-derived He will be enhanced. Thus, the ³He/⁴He ratios of marine ferromanganese nodules may shed light on their genetic relation.

We have measured element and isotope compositions of He

Table 1 Element and isotope compositions of light noble gas in marine ferromanganese nodules

No.	Code	Location (° N ° E)		Depth (m)	⁴ He (mm ³ STP g ⁻¹)	²⁰ Ne (mm ³ STP g ⁻¹)	³⁶ Ar (mm ³ STP g ⁻¹)	³ He/ ⁴ He (×10 ⁻⁶)	⁴⁰ Ar/ ³⁶ Ar	³⁸ Ar/ ³⁶ Ar	⁴⁰ Ar _{rad} (mm ³ STP g ⁻¹)
Mariana Trough											
1	KH-84-1-16-001	20 14	143 53	3,400-3,580	1.3 × 10 ⁻⁷	1.3 × 10 ⁻⁹	2.2 × 10 ⁻⁹	4.4 ± 1.0	835 ± 6	0.186 ± 0.001	1.2 × 10 ⁻⁶
2	KH-84-1-19-A	20 13	144 04	2,730-2,960	1.3 × 10 ⁻⁷	8.2 × 10 ⁻¹⁰	2.2 × 10 ⁻⁹	15.1 ± 1.6	807 ± 16	0.187 ± 0.002	1.2 × 10 ⁻⁶
3	KH-84-1-19-B	20 14	144 05		9.4 × 10 ⁻⁸	<3 × 10 ⁻⁸	5.0 × 10 ⁻⁸	4.3 ± 1.8	316 ± 4	0.189 ± 0.001	1.0 × 10 ⁻⁶
4	KH-84-1-19-C	20 14	143 05		—	—	4.0 × 10 ⁻⁸	—	333 ± 4	0.188 ± 0.002	1.5 × 10 ⁻⁶
5	KH-84-1-25	20 19	143 43	2,810-2,910	4.9 × 10 ⁻⁸	3.8 × 10 ⁻⁹	4.3 × 10 ⁻⁸	22.5 ± 1.7	324 ± 6	0.188 ± 0.001	1.2 × 10 ⁻⁶
Ogasawara region											
6	KH-84-1-2-124	26 16	142 58	1,430-1,570	2.6 × 10 ⁻⁸	<1 × 10 ⁻⁸	7.0 × 10 ⁻⁸	1.5 ± 0.6	311 ± 3	0.184 ± 0.003	1.1 × 10 ⁻⁶
7	KH-84-1-3-606	26 04	144 19	1,050-1,290	9.8 × 10 ⁻⁹	2.5 × 10 ⁻¹⁰	2.9 × 10 ⁻⁹	59.3 ± 9.4	384 ± 2	0.185 ± 0.001	2.6 × 10 ⁻⁷
8	KH-84-1-3-607	26 04	144 19		5.8 × 10 ⁻⁸	4.0 × 10 ⁻⁹	1.2 × 10 ⁻⁸	5.9 ± 1.5	290 ± 4	0.188 ± 0.001	—
9	KH-84-1-5-33	26 13	144 05	2,300-2,430	1.2 × 10 ⁻⁷	<1 × 10 ⁻⁸	2.7 × 10 ⁻⁸	10.2 ± 1.0	313 ± 6	0.188 ± 0.001	4.7 × 10 ⁻⁷
10	KH-84-1-27-017	26 12	144 11	840-890	1.2 × 10 ⁻⁸	<1 × 10 ⁻¹⁰	1.7 × 10 ⁻⁹	6.0 ± 5.0	432 ± 1	0.185 ± 0.002	2.3 × 10 ⁻⁷

Error assigned to the isotope ratios is 1 s.d.

and Ar gases in 10 nodules dredged from Mariana Trough (opening back-arc basin) and the Ogasawara region on the Hakuho Maru Cruise (KH-84-1¹¹; Fig. 1). After being washed with distilled water, samples were dried and preheated *in vacuo* at 150 °C in a gas extraction system for two days to reduce surface air contamination. Each sample was placed in a double-walled tantalum crucible with a surrounding tantalum heater and extracted gases at 1,700 °C (refs 12, 13). Gas purification was carried out by the two-stage Ti/Zr getter. The He-Ne and Ar fractions were separated by use of charcoal traps held at -196 °C. The noble gas isotopes were analysed by a single focusing mass spectrometer (6-60-SGA, Nuclide Co.). A resolving power of ~600 at 1% peak height was attained to separate ³He from H₃ + HD (ref. 14). The measured ³He/⁴He ratios were calibrated against the ratio of air standard. Element abundances of Ne and Ar and isotope compositions of Ar were also determined under the same conditions. Possible interference of hydrocarbon with ³⁶Ar and ³⁸Ar measurements was eliminated by the high resolving power. The measuring system, including the gas extraction and purification line and the mass spectrometer, has not yet been used in the analysis of extraterrestrial materials to eliminate the risk of 'memory' effect. Observed noble gas data are listed in Table 1. All tabulated numbers are blank-corrected.

The ³He/⁴He ratios of nodules vary significantly from 1.5 × 10⁻⁶ to 59.2 × 10⁻⁶. Every sample has a higher ³He/⁴He ratio than the atmospheric level. The highest value, 59.3 × 10⁻⁶ (sample 7), is extraordinarily high when compared with terrestrial materials and is almost equivalent to deep mantle components from the Hawaiian hot spot region (plume-type He)¹⁵. There is no evidence of direct supply of deep mantle material in the Ogasawara region where the nodule was recovered. Radiogenic ³He, produced by nuclear reactions¹⁶ (⁶Li(n, α)³He) cannot account for such a high ³He/⁴He ratio because of the dilution of radiogenic ⁴He by α decay of U and Th. Fallout from nuclear explosions may also be excluded as nodules generally grow at an extremely low rate on the deep sea floor. Furthermore, the surface of the samples was carefully removed before noble gas analysis. The only terrestrial materials known to have such a high ³He/⁴He ratio as we observed¹⁷ is diamond. This suggests an extraterrestrial origin of He.

The second highest ratio, 22.5 × 10⁻⁶ (sample 5), is apparently higher than that of basalt glasses dredged from the same region, Mariana Trough¹⁸, with a value of 11.6 × 10⁻⁶. There is no evidence of discharge of plume-type He in the trough or of other processes that increase the ³He/⁴He ratio sufficiently (see above), pointing to an extraterrestrial origin for nodules from both Mariana Trough and the Ogasawara region. The nodules with relatively lower ³He/⁴He ratios may have arisen by either heterogeneous conglomeration of extraterrestrial materials or dilution of *in situ* radiogenic ⁴He after the inventory.

In 1964, Merrihue reported extremely high ³He/⁴He ratios of up to 1 × 10⁻⁴ in ferromagnetic separates from Pacific deep ocean red clay¹⁹. The high ³He/⁴He ratio was explained by the addition

of extraterrestrial debris in the sediments. A subsequent study confirmed that higher ³He/⁴He ratios are common to pelagic sediments whose sedimentation rate is low²⁰. A recent study reported that homogeneous fallout of stratospheric dust into the sediments can account for the ³He/⁴He ratio anomaly²¹. A close relation between the occurrence of the marine ferromanganese nodules and ocean red clay was established by Cronan²², who found that nodules do not grow on the ocean floor in places where the sedimentation rate is higher than 7 mm per 1,000 yr. Thus, the sedimentation rate is closely linked to both ³He/⁴He ratios in the sediments and the occurrence of the nodules, which suggests similar mechanism of genesis for nodules with high ³He/⁴He ratios and such sediments.

The presence of cosmic spherules was first recognized in manganese nodules by Murray and Renard²³. A subsequent study revealed their characteristics using techniques of X-ray diffraction and electron microprobe analysis²⁴. They were inferred to be two types of particles, one derived from iron meteorites and the other from stony meteorites. These particles are extremely small, with diameters of 100–200 μm. Materials with a composition similar to these particles may account for the observed high ³He/⁴He ratios in manganese nodules, since the

Table 2 Major and minor elements of marine ferromanganese nodules

Element	1 KH-84-1-16-001	2 KH-84-1-19-A	7 KH-84-1-3-606	10 KH-84-1-27-017
Major (%)				
MnO ₂	22.4	17.3	45.9	40.9
Fe ₂ O ₃	25.5	25.2	13.1	25.4
SiO ₂	19.7	23.3	9.51	4.70
Al ₂ O ₃	7.18	7.90	3.26	1.28
TiO ₂	1.30	1.40	1.24	1.36
Na ₂ O	1.82	2.15	2.51	1.72
K ₂ O	1.16	0.93	1.40	0.48
CaO	3.82	3.93	4.32	3.57
MgO	2.21	2.17	2.08	1.89
P ₂ O ₅	0.96	0.94	1.39	1.32
H ₂ O (+)	10.8	9.64	11.4	12.2
Minor (p.p.m.)				
Ba	2,060	1,750	2,460	3,110
Sr	1,290	1,180	1,380	2,080
Li	31	22	110	16
V	460	430	800	920
Co	600	540	5,570	6,510
Ni	1,290	750	7,500	4,090
Cu	810	720	1,100	540
Zn	490	470	1,240	660
Y	160	430	100	210
La	180	170	100	400
As	140	120	140	350
Sb	9.1	9.2	40	48
Th	39	8.9	21	34
Sc	21	22	7	10
W	30	20	160	250
Mo	70	40	820	750
Hf	7.1	6.5	5.1	8.5

$^3\text{He}/^4\text{He}$ ratio of meteorites is known to be extremely high, up to 0.26 (refs 25, 26). Ozima *et al.*²¹ suggested that particles smaller than 10 μm in diameter would be carriers of the high $^3\text{He}/^4\text{He}$ ratio, applying the calculation on the loss from particles at atmospheric impact heating²⁷. The identity of the He carrier remains to be determined.

Chemical compositions of major and trace elements of representative nodules were determined by means of inductively coupled argon plasma optical emission spectroscopy and instrumental neutron activation analysis. Preliminary results are given in Table 2. Based on the Mn/Fe ratios, nodules from Mariana Trough (samples 1 and 2) are largely of hydrogenetic type, whereas those from the Ogasawara region are diagenetic (sample 7) and intermediate (sample 10) types. Trace elements such as Co, Sb and Ni are relatively depleted in Mariana nodules compared to Ogasawara nodules, which agrees well with the respective Mn/Fe ratios and genetic types.

There is no relation between the $^3\text{He}/^4\text{He}$ ratios and the chemical compositions of nodules, which implies that differ-

ences in genetic conditions, such as those between hydrogenetic and diagenetic types, cannot affect the $^3\text{He}/^4\text{He}$ ratio. No apparent relation is found between elemental abundances of ^4He , ^{20}Ne and ^{36}Ar and chemical compositions of nodules. In contrast to other noble gas data, contents of radiogenic ^{40}Ar are higher in Mariana samples than in Ogasawara samples. It was confirmed that diagenetic type nodules grow faster than hydrogenetic type²⁸. Abundances of radiogenic ^{40}Ar in nodules show agreement with the reported regularity and the genetic signature arrived at by chemical composition analysis.

In conclusion, the high $^3\text{He}/^4\text{He}$ ratios in marine ferromanganese nodules cannot be attributed to terrestrial materials, but strongly suggest the occurrence of extraterrestrial debris in the nodules.

We thank Professor K. Kobayashi of the University of Tokyo for arranging sample collection through the KH-84-1 Cruise of the Research Vessel of the University of Tokyo. We also thank Professor M. Ozima and Dr Y. Nakamura for providing valuable comments and help in experimental procedures.

Received 25 March; accepted 2 August 1985.

1. Calvert, S. E. & Price, N. B. *Mar. Chem.* **5**, 43-74 (1977).
2. Lyle, M., Dymond, J. & Heath, G. R. *Earth planet. Sci. Lett.* **35**, 55-64 (1977).
3. Halback, P., Heibisch, U. & Scherhag, C. *Chem. Geol.* **34**, 3-17 (1981).
4. Harriss, R. C. *Nature* **219**, 54-55 (1968).
5. Ozima, M. & Podosek, F. A. in *Noble Gas Geochemistry* (Cambridge University Press, 1983).
6. Mamyrin, B. A. & Tolstikhin, I. N. in *Helium Isotopes in Nature* (Elsevier, Amsterdam, 1984).
7. Clarke, W. B., Horowitz, R. M. & Broecker, W. S. *Init. Rep. DSDP* **15**, 777-781 (1973).
8. Sano, Y. & Wakita, H. *Init. Rep. DSDP* (in the press).
9. Wakita, H., Sano, Y., Fujii, N. & Takeuchi, A. *Init. Rep. DSDP* (in the press).
10. Lupton, J. E. & Craig, H. *Science* **214**, 73-86 (1981).
11. *Preliminary Report of The Hakuhō Maru Cruise KH-84-1* (Ocean Research Institute, University of Tokyo, 1985).
12. Takaoka, N. *Mass Spectrosc.* **24**, 73-86 (1976).
13. Ozima, M. & Zashu, S. *Earth planet. Sci. Lett.* **62**, 24-40 (1983).
14. Sano, Y., Tominaga, T., Nakamura, Y. & Wakita, H. *Geochim. J.* **16**, 237-245 (1982).

15. Kaneoka, I. & Takaoka, N. *Earth planet. Sci. Lett.* **39**, 382-386 (1978).
16. Morrison, P. & Pine, J. *Ann. N.Y. Acad. Sci.* **62**, 71-79 (1955).
17. Ozima, M. & Zashu, S. *Science* **219**, 1067-1068 (1983).
18. Poreda, R. & Craig, H. *EOS* **60**, 969 (1979).
19. Merrihue, C. *Ann. N.Y. Acad. Sci.* **119**, 351-367 (1964).
20. Krylov, A. Yu., Mamyrin, B. A., Silin, Yu. I. & Khabarin, L. V. *Geochemistry Int.* **10**, 202-205 (1973).
21. Ozima, M., Takayanagi, M., Zashu, S. & Amari, S. *Nature* **311**, 448 (1984).
22. Cronan, D. S. in *Marine Manganese Deposits* (ed. Glasby, G. P.) 11-44 (Elsevier, Amsterdam, 1977).
23. Murray, J. & Renard, A. *Proc. R. Soc. Edinb.* **12**, 474-495 (1884).
24. Finkelman, R. B. *Science* **167**, 982-984 (1970).
25. Schultz, L., Funk, H., Nyquist, L. & Signer, P. *Geochim. cosmochim. Acta* **35**, 77-88 (1971).
26. Levisky, L. K. & Komarov, A. N. *Geochim. cosmochim. Acta* **39**, 275-284 (1975).
27. Rajan, R. S. *et al. Nature* **267**, 133-134 (1977).
28. Reyss, J. L., Marchig, V. & Ku, T. L. *Nature* **295**, 401-403 (1982).

Search for the origin of exotic helium in deep-sea sediments

Sachiko Amari & Minoru Ozima

Geophysical Institute, University of Tokyo, Tokyo 113, Japan

Merrihue¹ first observed helium with a very high $^3\text{He}/^4\text{He}$ isotope ratio ($\geq 10^{-4}$; here termed exotic helium) in a magnetic separate of a deep-ocean sediment and attributed it to input of extraterrestrial material. Krylov *et al.*² subsequently reported the existence of exotic He in several deep-sea sediments, while Tilles³ measured argon isotope ratios in magnetic separates of two ocean sediments. Observing that the $^3\text{He}/^4\text{He}$ ratio is inversely correlated with the sedimentation rate, Ozima *et al.*⁴ concluded that the exotic He is carried by interplanetary dust particles (IDPs). We have now examined other deep-sea sediments to characterize the exotic He and to identify its carrier, and have confirmed that the exotic He resides mainly in a strongly magnetic fraction. Stepwise degassing experiments suggest that the exotic He consists of a single component. The carrier of the He is likely to be chondritic IDPs. The higher level of exotic He in magnetic fractions of sediments can be explained by secondarily produced magnetite in IDPs as a result of thermal shock during atmospheric entry.

We studied deep-sea pelagic sediments from three different sites: (1) Piston-cored sediment⁵, KH75-3-5-2 (26°00.8' N, 150°00.0' E; uncorrected depth, 5,850 m; core length, 907 cm); (2) box-cored, siliceous fossil-rich clay⁶, GH79-1-1476 (9°49.96' N, 167°17.68' W; corrected depth, 5,151 m); and (3) box-cored, very dark brown siliceous and calcareous fossil-rich mud⁷, GH80-1-1627 (5°27.32' S, 163°46.01' W; corrected depth, 4,995 m). A hand magnet was used to separate magnetic fractions from the pulverized samples, which were suspended in distilled water.

In the case of KH75-3-5-2 (662-711-cm interval), three different grades of magnetic fractions were separated according

to the intensity of their magnetization; these were designated M1, M2 and M3, where the smallest number corresponds to the strongest magnetization. Their weight fractions were ~0.22, 0.09 and 0.27%, respectively. Compared with M2 and M3, M1 was relatively abundant in particles <10 μm .

The two box-cored samples (GH79-1-1476, GH80-1-1627) were separated only into magnetic and non-magnetic fractions. In the case of the magnetic fraction of GH80-1-1627, most of the particles were a few micrometres in size, but the particles in the magnetic fraction of GH79-1-1476 ranged from 50 to 100 μm , with few particles smaller than 10 μm .

Both the highest $^3\text{He}/^4\text{He}$ ratio and the highest ^3He content were observed in M1, the most strongly magnetic fraction of KH75-3-5-2. The magnetic fractions of the box-cores also had higher $^3\text{He}/^4\text{He}$ ratios and ^3He contents (Table 1) than the bulk sediments. In only 0.22 wt% of the bulk sediment, M1 contained ~30% of bulk ^3He , with a ^3He content of $\sim 3.6 \times 10^{-9} \text{ cm}^3 \text{ STP per g}$. However, the non-magnetic fraction also showed a higher $^3\text{He}/^4\text{He}$ ratio and He content than common terrestrial materials. This is discussed below.

The $^3\text{He}/^{20}\text{Ne}$ ratios were 0.03-0.07, which is similar to the planetary ratio (0.03)⁸ but slightly smaller than the solar wind He implanted in lunar soils (0.09)⁸. As most of the observed $^3\text{He}/^4\text{He}$ ratios (2×10^{-4}) differ from both the planetary (1.4×10^{-4}) and the solar-type He (4×10^{-4}), we first suspected it to be a mixture of several components, such as radiogenic He and/or spallogenic He, in addition to the two principal components. However, when M1 was subjected to stepwise heating, the $^3\text{He}/^4\text{He}$ ratios were almost the same (2×10^{-4}) for most of the temperature fractions (Table 1). Figure 1a shows the thermal release pattern of ^4He , while Fig. 1b shows a ^3He - ^4He correlation diagram. The strong linear correlation ($^3\text{He}/^4\text{He} = 2 \times 10^{-4}$) suggests that He in M1 comprises a single component. This is supported by the Arrhenius plot in Fig. 2, which shows a straight line yielding a single activation energy of $E = 17 \text{ kcal mol}^{-1}$; different components would be more likely to be characterized by different outgassing mechanisms or by different activation energies. Solar wind He ($^3\text{He}/^4\text{He} = 4 \times 10^{-4}$) implantation

Table 1 Properties of sediment samples investigated

Sample	Weight (mg)	$^3\text{He}/^4\text{He}$ (10^{-4})	^3He (10^{-10} cm ³ STP per g)	^4He (10^{-6} cm ³ STP per g)	^{20}Ne (10^{-8} cm ³ STP per g)
KH-75-3-5-2 Interval: 662–711 cm					
Bulk	101.6				
Total		0.659 ± 0.132	0.258 ± 0.085	0.392 ± 0.050	
NM	104.5				
Total		0.516 ± 0.085	0.208 ± 0.058	0.402 ± 0.047	
M1	42.25				
Total		2.18 ± 0.24	36.10 ± 6.16	16.55 ± 0.98	5.18 ± 0.61
M1	55.85				
520 °C		1.99 ± 0.13	7.72 ± 1.28	3.88 ± 0.40	
690 °C		2.05 ± 0.11	10.74 ± 1.62	5.25 ± 0.51	
770 °C		1.91 ± 0.14	8.11 ± 1.51	4.25 ± 0.49	
860 °C		1.90 ± 0.13	5.71 ± 1.04	3.01 ± 0.34	
950 °C		1.76 ± 0.15	3.37 ± 0.73	1.91 ± 0.25	
1,030 °C		1.12 ± 0.46	0.321 ± 0.266	0.285 ± 0.120	
1,110 °C		3.59	0.185	0.141*	
M2	26.3				
Total		1.87 ± 0.32	5.78 ± 1.68	3.10 ± 0.37	
M3	55.5				
Total		1.75 ± 0.28	2.15 ± 0.57	1.23 ± 0.13	
GH79-1-1476					
M	41.91				
Total		1.38 ± 0.07	16.21 ± 2.52	11.79 ± 1.27	5.97 ± 3.17
GH180-1-1627					
M	30.53				
Total		2.37 ± 0.08	107.3 ± 13.4	45.26 ± 4.07	43.50 ± 8.55

NM, non-magnetic fraction; M, magnetic fraction; M1, M2 and M3, see text. Errors represent 1σ including the analytical error in the mass spectrometry and 50% of the hot blank.

* These values give only upper limits because of the very small amount of the extracted gas.

seems to be the only plausible mechanism to account for the large amount of trapped exotic He (^3He up to 1×10^{-8} cm³ STP per g). However, it is difficult to explain the finding that this ratio is lower than that of solar He in terms of the addition of radiogenic ^4He , as the exotic He is probably of a single component as discussed above. If we assume that ~99% of the original He was lost, the solar wind ratio can be reduced to the observed value. However, because the $^3\text{He}/^{20}\text{Ne}$ ratio is not very different from the value of solar wind-implanted lunar soil, such large fractionation seems unlikely, as the process that resulted in the isotope fractionation would have produced much larger elemental fractionation.

Thermomagnetic analyses of M1 and the magnetic fraction of GH80-1-1627 (GH80M) in a vacuum of $\sim 10^{-3}$ Pa at a magnetic field of $H = 0.45$ T, where H is the magnetic field in Teslas, produced almost reversible thermomagnetic curves on heating and cooling (Fig. 3). The principal Curie temperatures were ~ 540 °C for M1 and 575 °C for GH80M, with respective magnetizations of $39 \text{ Am}^2 \text{ kg}^{-1}$ and $7 \text{ Am}^2 \text{ kg}^{-1}$. Based on the thermomagnetic curves and magnetic intensities as well as on chemical analyses by atomic absorption spectroscopy (M1: Fe, 31%; Mn, 0.4%; Ni, 0.4%; GH80M: Fe, 14%; Mn, 0.4%; Ni, 300 p.p.m.), we conclude that the ferromagnetic separates are essentially magnetite (with a small amount of ulvöspinel in M1). Because magnetite in deep-sea sediments is commonly nearly stoichiometric⁹, most of the ferromagnetic grains may be of terrestrial origin, in which case the thermomagnetic curve for the carrier of exotic He is likely to be completely masked. If we assume that IDPs are the He carrier, we can use the ^4He content observed in IDPs ($\sim 0.1 \text{ cm}^3 \text{ STP per g}$)¹⁰ to estimate the amount of IDPs in the magnetic fractions; incorporation of ~ 100 p.p.m. IDPs in the magnetic fractions would be enough to account for the observed He content.

The exotic He has several characteristic features. First, it must be quite resistant to seawater weathering. Chondritic IDPs are coated with amorphous materials of low atomic number^{11,12} which are considered to be carbonaceous matter. Perhaps this

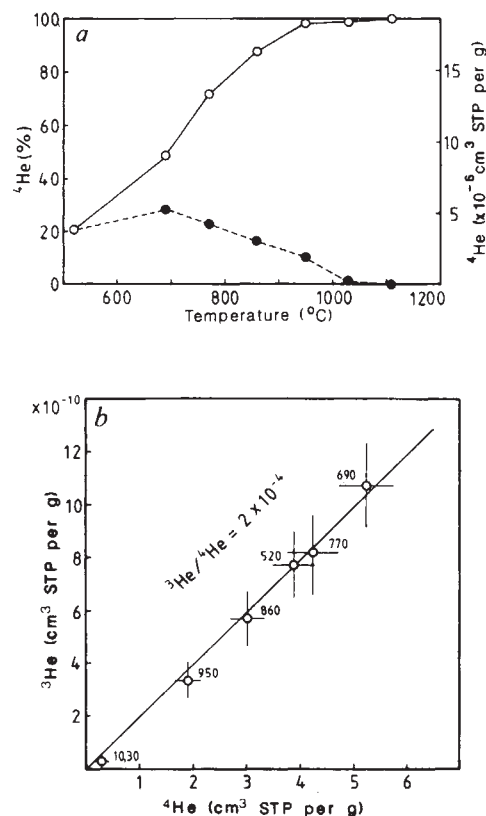


Fig. 1 *a*, Helium thermal release pattern. $\circ$, Cumulative value; $\bullet$, value for each temperature step. *b*, $^3\text{He}/^4\text{He}$ correlation diagram. The amount of ^3He is plotted against the amount of ^4He . Numbers next to each data point indicate temperature (°C). Errors bars are 1σ .

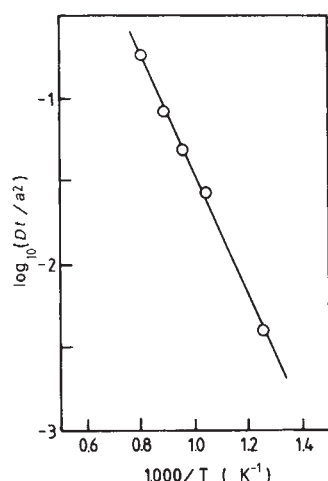


Fig. 2 Arrhenius plot for stepwise heating. The slope yields an activation energy of 17 kcal mol⁻¹. D , diffusion coefficient; t , time interval for each temperature step (1 h); a , radius of a particle.

coating gives good protection against seawater weathering. Second, the stepwise heating experiment shows that He is trapped fairly tightly; a 950 °C temperature fraction after four steps still contains ~10% of the total He (Fig. 1a). Extrapolation of the experimental data on the Arrhenius plot to the ambient temperature in the deep-sea bottom (~2 °C) gives $D/a^2 = 2.3 \times 10^{-15} \text{ s}^{-1}$ (where D is the diffusion coefficient and a the radius of a particle), which suggests that ~40% of He can still remain in the magnetic fraction after half a million years. Third, the high $^3\text{He}/^4\text{He}$ ratio is highly concentrated in the magnetic fraction. As discussed previously^{1,2,4}, IDPs seem to be the most plausible candidate for the exotic He carrier in deep-sea sediments. Most IDPs have an elemental abundance similar to that of CI chondrites¹³, which contain magnetite of ~8 wt% (ref. 14). However, after atmospheric entry, IDPs would contain secondarily produced magnetite; X-ray diffraction analyses by Fraundorf *et al.*¹⁵ revealed a correlation between the strength of magnetite spots and other signs of heating and showed that the least altered IDPs had weak patterns, suggesting that in the most primitive IDPs, non-crystallized materials are the major phase. These workers concluded that finely divided magnetite and olivine are produced by decomposition of poorly ordered silicate materials. The amorphous silicates are metastable and would therefore easily change into stable minerals, that is, magnetite and olivine, following a small thermal stimulus, without causing large He loss. For example, in the case of IDPs with 3 g cm⁻³ and 10 µm size (a typical size in stratospheric particles¹⁶), ~10% of the total He would be lost during atmospheric entry, as estimated from the data for the maximum temperature reached (~1,000 °C) without causing melting of the IDPs¹⁷, the duration of heating (~10 s)^{18,19} and the results of stepwise heating. Therefore, most chondritic IDPs would contain magnetite originating from the thermal shock of atmospheric entry. In fact, it has been reported that most X-ray diffraction patterns of chondritic IDPs indicate the presence of magnetite and pyrrhotite²⁰. Hence, the amount of magnetite must be larger than that inferred from CI chondrites. In deep-sea sediments, most terrestrial particles are non-magnetic, whereas magnetic IDPs must be more abundant than the non-magnetic IDPs. This would explain why the exotic He resides essentially in the magnetic fractions.

We thank Professor K. Yamakoshi and Dr K. Fujioka for providing the samples, Dr T. Furuta for help with the thermomagnetic analyses and Dr M. Imamura for help with the atomic absorption analyses. S.A. thanks Professors K. Ito and J. Matsuda for their continuous encouragement. S.A. is on leave of absence from the Division of Science of Materials, Graduate School of Science and Technology, Kobe University, Nada, Kobe 657, Japan.

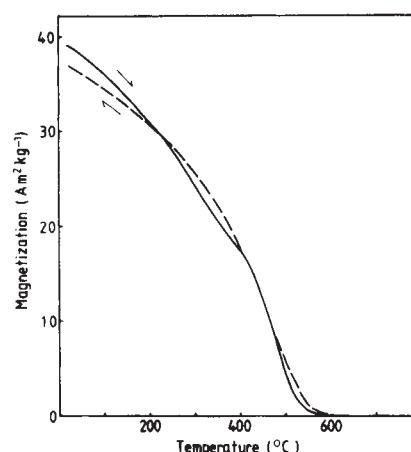


Fig. 3 Thermomagnetic analysis of M1. Solid line, heating curve; dashed line, cooling curve.

Received 15 May; accepted 21 August 1985.

1. Merrihue, C. *Ann. N.Y. Acad. Sci.* **119**, 351-367 (1964).
2. Krylov, A. Ya., Mamyrin, B. A., Silin, Yu. I. & Khabarin, L. V. *Geochem. Int.* **10**, 202-205 (1973).
3. Tilles, D. *Science* **153**, 981-984 (1966).
4. Ozima, M., Takayanagi, M., Zashu, S. & Amari, S. *Nature* **311**, 448-450 (1984).
5. Kobayashi, K., Tonouchi, S., Furuta, T. & Watanabe, M. *Bull. Ocean Res. Inst.* **13**, 116-121 (1980).
6. Mizuno, A. *et al.* in *Cruise Rep.* **15** (ed. Mizuno, A.) 1-31 (Geological Survey of Japan, Tsukuba, 1981).
7. Mizuno, A. *et al.* in *Cruise Rep.* **18** (eds Mizuno, A. & Nakao, S.) 1-35 (Geological Survey of Japan, Tsukuba, 1982).
8. Ozima, M. & Podosek, F. A. *Noble Gas Geochemistry*, 68-126 (Cambridge University Press, 1983).
9. Løvlie, R. *Earth planet. Sci. Lett.* **21**, 315-320 (1974).
10. Rajan, R. S. *et al.* *Nature* **267**, 133-134 (1977).
11. Fraundorf, P. & Shirk, J. *Proc. 10th lunar planet. Sci. Conf.* 951-976 (1979).
12. Fraundorf, P. *Geochim. cosmochim. Acta.* **45**, 915-943 (1981).
13. Brownlee, D. E., Olszewski, E. & Wheelock, M. *Lunar planet. Sci.* **13**, 71-72 (1982).
14. Hyman, M., Rowe, M. W. & Herndon, J. M. *Geochem. J.* **13**, 37-39 (1979).
15. Fraundorf, P., Brownlee, D. E. & Walker, R. M. in *Comets* (ed. Wilkening, L. L.) 383-409 (University of Arizona Press, 1982).
16. Brownlee, D. E. *Rev. Geophys. Space Phys.* **17**, 1735-1743 (1979).
17. Fraundorf, P. *Geophys. Res. Lett.* **10**, 765-768 (1980).
18. Brownlee, D. E., Tomandl, D. A. & Olszewski, E. *Proc. 8th lunar Sci. Conf.* 149-160 (1977).
19. Brownlee, D. E. *Ann. Rev. Earth planet. Sci.* **13**, 147-173 (1985).
20. Brownlee, D. E. in *Protostars and Planets* (ed. Gehrels, T.) 134-150 (University of Arizona Press, 1978).

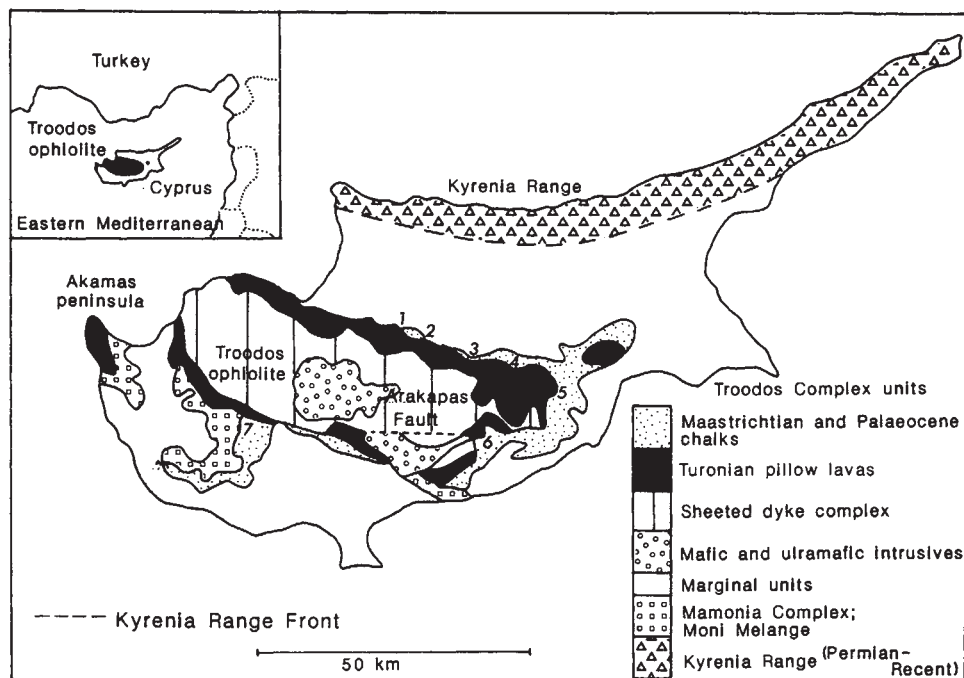
Palaeorotation of the Troodos microplate, Cyprus

T. M. M. Clube*, K. M. Creer* & A. H. F. Robertson†

Departments of * Geophysics and † Geology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JW, UK

The Troodos ophiolite represents one of the best preserved fragments of ocean-floor crust that is exposed on land. As the extrusive series of the ophiolite retains a stable magnetization that is directed westward, the ocean crust is considered to have rotated anticlockwise through 90° since its formation in the Late Cretaceous¹. To determine the timing of the rotation event, over 3,000 orientated samples have been collected from the ophiolite and from its *in situ* Turonian to Recent sedimentary cover. Here we report that pelagic chalks immediately overlying the highest extrusive lava units of the ophiolite complex give reliable magnetic vector directions indicating that at least 60° of rotation occurred before the Lower Eocene and that rotation was completed by the end of the Lower Eocene. Rotation cannot have occurred within a single late Miocene event, as reported previously². Regional geological considerations support the rotation of only a small fragment of oceanic crust that was stranded adjacent to an active continental margin. In this tectonic setting the oblique consumption of oceanic crust beneath Troodos crust could provide the necessary driving force for tectonic rotation.

Fig. 1 Simplified geological map of Cyprus (with inset showing the location of Cyprus in the eastern Mediterranean area). 1-7, Sampling sites within Maastrichtian and Palaeocene chinks.



The Troodos ophiolite complex is a preserved remnant of late Cretaceous oceanic lithosphere which retains an intact, largely undeformed, lithostratigraphy³. The characteristic oceanic crustal layers now crop out concentrically about the focus of relatively recent uplift (Fig. 1). The presence of a well-developed sheeted dyke complex constitutes unequivocal evidence that crustal accretion occurred at a linear spreading axis orientated parallel to the predominant north-south dyke trend^{4,5}.

Extrusive pillow lavas crop out around the periphery of the main igneous complex and on a sliver of ophiolitic basement exposed on the Akamas Peninsula, north-west Cyprus (Fig. 1). After alternating field (AF) magnetic cleaning, in fields of <10 mT, samples collected from lavas along the northern Troodos margin clearly recorded stable westerly declinations, consistent with previous studies¹ (Table 1). The samples are of normal polarity, implying that the crust rotated 90° anticlockwise after formation. The original spreading axis was therefore orientated east-west. In detail, a comparison of structurally well-constrained magnetic declinations from numerous lava outcrops shows that intracrustal block-rotations in the horizontal plane have been considerable, especially in the vicinity of major vertical fracture zones (for example, Arakapas fault belt⁶). A complementary study⁷ has revealed that bulk-rotation in the vertical plane on listric faults was an important control on relative dyke attitude. Such rotations are common in active extensional tectonic regimes.

The age of the oceanic crust itself (91–88 Myr⁸) provides a lower limit on the timing of tectonic rotation. In order to constrain the event further, the palaeomagnetic study was extended

into the *in situ* sedimentary cover of the ophiolite. Typically, the stratigraphically highest lavas are overlain by umbers (ferromanganiferous oxide sediments, dated as Turonian⁸) and radiolarites of late Campanian age⁹, in turn passing into an unbroken succession (50–1,500 m) dominated by pelagic carbonates of Maastrichtian to early Miocene age¹⁰.

Following standard collection and laboratory procedures, the remanent magnetization of samples was measured with a two-axis cryogenic magnetometer. The detailed thermal and AF demagnetization behaviour was studied on pairs of two-component plots (Fig. 2a, b). Samples from the Maastrichtian and Palaeocene chinks around the Troodos periphery characteristically record a generally higher level of magnetization (10–100 $\mu\text{A m}^{-1}$) than Eocene and Oligocene chinks, which often only retain a remanent magnetization intensity comparable with the sensitivity of the magnetometer (4 $\mu\text{A m}^{-1}$). Accuracy in repeatability during stepwise demagnetization was used as a rejection criterion¹¹.

The remanent magnetizations of Maastrichtian and Palaeocene chinks typically include two distinct components, each commonly exhibiting partially overlapping coercivity spectra. A low coercivity present-day field component is removed in low alternating fields (15 mT). A harder component, presumably primary, has unblocking temperatures between 500 °C and 600 °C and moderate coercivities. The 'soft' component is probably carried exclusively by multi-domain magnetite, while the primary component seems to be carried by a combination of single-domain titanium-poor magnetite and specular haematite (Fig. 2d).

The characteristic cleaned remanent directions were corrected for simple tectonic tilt, then plotted on a stereographic projection. Diametrically opposite normal and reversed polarities are clearly represented (Fig. 2c). After inversion of reversed-polarity directions through the origin, a mean remanence direction is calculated to be declination (dec.) = 334, inclination (inc.) = +23, α_{95} = 17.3, N = 217 compared with dec. = 274, inc. = +36, α_{95} = 12.3, N = 663 for the underlying ophiolitic extrusive (Table 1). Preliminary results from the metalliferous umbers and radiolarites show declinations similar to those of the ophiolite complex. Anomalously low inclinations recorded in the sediments, and at occasional intervals in the overlying chinks, are probably due to the rotation of detrital magnetic carriers during sedimentary compaction, a process which leaves declinations unaffected¹².

Table 1 Comparative mean cleaned magnetic directions of ophiolite and its sedimentary cover after AF demagnetization

Rock age	N	n	dec.	inc.	α_{95}
TR Turonian	11	663	274	+36	12.3
PE Maastrichtian-Palaeocene (N)	5	116	336	+32	13.2
PE Maastrichtian-Palaeocene (R)	6	101	152	-14	15.3
PE Lower Eocene	5	136	357	+38	10.1

TR, Troodos pillow lavas (north Troodos and Akamas); PE, pelagic chinks. N , number of sites; n , number of samples; α_{95} = 95% radius of cone of confidence; N, normal polarity; R, reversed polarity. The declination values constrain ~60° of the rotation prior to the end of the Palaeocene and a further 30° during the Lower Eocene.

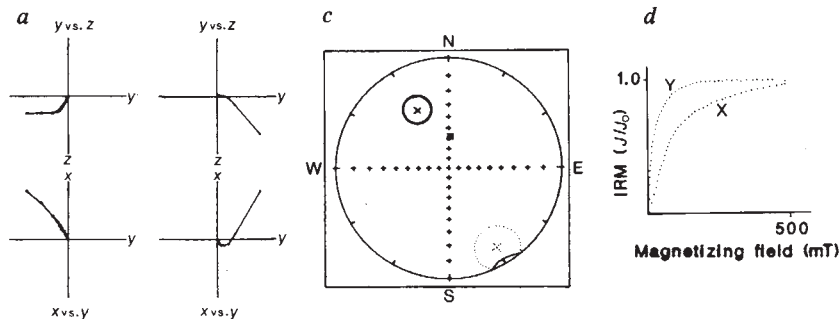


Fig. 2 Summary of magnetic properties of Maastrichtian and Palaeocene chalks. *a*, *b*, AF demagnetization plots of normal and reversed-polarity samples, respectively (4-mT incremental steps). Spurious acquired remanent magnetizations at high fields (>50 mT) prevented further AF treatment; thermal demagnetization was then preferred. *c*, Stereographic projection showing both normal (full circle) and reversed (dotted) polarities. *d*, Isothermal remanent magnetization (IRM) acquisition curves showing the likely presence of magnetic and specular haematite in some samples (X), with magnetite the probable sole carrier in other samples (Y). *J* represents induced magnetization in external field. J_0 is induced magnetization at 500 mT.

Magnetic directions in the sediment cover must reflect motion of the underlying ophiolitic basement. Thus, the declination values indicate relatively rapid rotation of Troodos oceanic crust between Turonian and early Eocene time, over ~30 Myr (see Table 1). Declinations recorded in the Maastrichtian sediments show that much of the rotation took place in Campanian-Maastrichtian time, a period of only a few million years, immediately following the end of spreading-ridge volcanism.

The geological field evidence suggests strongly that there were major tectonic boundaries close to the Troodos sea floor at the time of rotation and it may be appropriate to consider rotating crust no larger than Cyprus itself. Certainly on a larger scale it is clear that the existing palaeomagnetic data from southern Turkey and Africa^{13,14} indicate that these areas have not rotated to anything like the same extent as the Troodos ophiolite.

In south-west Cyprus, quite coarse-grained terrigenous sediments (Kannaviou Formation) appear interbedded with ridge-related manganese radiolarites (Perapedhi Formation, of Turonian-early Maastrichtian age), only several metres stratigraphically above the uppermost Troodos lavas¹⁵. In addition, Mesozoic continental margin lithologies are present as detached blocks (olistoliths) within the *in situ* sedimentary cover of the Troodos ophiolite in southern Cyprus¹⁶ (Moni Melange, Fig. 1). These observations require the Troodos ophiolite to have been located close to a Mesozoic continental margin as early as Turonian-early Maastrichtian time (only ~5–10 Myr after its genesis). Other remnants of this continental margin are now preserved in the Mamonia Complex (south-west Cyprus)^{17,18} and the Kyrenia Range (north Cyprus)¹⁹. Structural studies in these areas confirm that important strike-slip deformation occurred contemporaneously with rotation of the Troodos ophiolite^{19,20}.

It would thus appear that the tectonic rotation involved interaction of a small oceanic crustal segment with a Mesozoic continental margin. Do any analogues of such a tectonic juxtaposition exist? One area of immediate interest is off the western margin of the North America plate, where subduction has carried the Farallon plate obliquely against the consuming margin of the California borderland²¹. As a result of transform-trench interactions, the original plate has disintegrated into a number of small fragments (sometimes <10⁴ km²) each recording considerable independent relative rotations (for example, the Gorda plate and Explorer plates²²). In this setting, a key control on the rotation of larger lithospheric plates like the Northern Cocos plate (Fig. 3a) appears to be related to the failure of young hot buoyant oceanic crust to be subducted, resulting in the spreading-ridge crust pivoting towards the continental margin at the locus of ridge-trench collision²³.

Rotations of small lithospheric units can also occur in the arc-trench gap area. Here, a substantial component of oblique subduction is known to give rise to marked lateral shear in the overriding plate and this can lead to rotations of terranes located behind the trench in a fore-arc setting^{24,25}. In the case of the Sunda arc, oblique convergence has led to the north-westerly displacement of a (>10⁵ km²) crustal wedge relative to the Indian Plate (Fig. 3b).

Our favoured model for the ophiolite palaeorotation involves oblique northward subduction of Tethyan oceanic crust, leading to impingement of the spreading ridge on a Mesozoic continental margin to the north²⁶ (Fig. 3c). To be consistent with recent geochemical evidence^{27,28}, which appears to confirm an origin of the Troodos ophiolite above a subduction zone, it seems probable that the ophiolite was formed in some type of 'fore-arc'²⁹ or 'pre-arc'³⁰ setting. Remnants of the Mesozoic continental margin with which the Troodos oceanic crust interacted are now preserved in south-west (Mamonia Complex) and north (Kyrenia Range) Cyprus.

Thus, we draw three main conclusions. First, the remarkable rotation of the Troodos ophiolite occurred relatively quickly in the Late Cretaceous-early Tertiary interval quite soon after its genesis. Second, the crustal area involved was relatively small, with at least one of its boundaries possibly being preserved within the present area of Cyprus. Third, although the exact mechanism of the rotation cannot yet be specified in detail, it seems likely that a major driving force resulted from the oblique subduction of oceanic crust beneath the Troodos ophiolite which was then sited in a 'fore-arc' setting immediately adjacent to a continental margin to the north, as shown in Fig. 3c. Additional palaeomagnetic and field studies are in progress to develop and test this model.

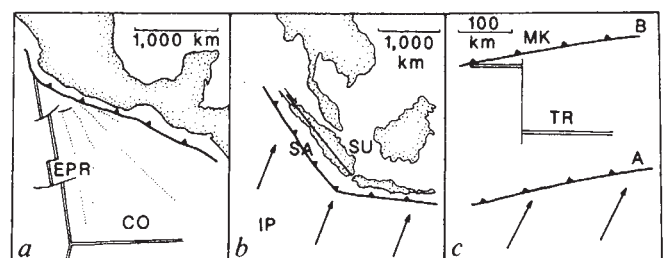


Fig. 3 Oceanic microplate rotations at active margins. ▲, Trench;, seafloor magnetic anomalies; =, spreading ridge; arrows indicate direction of convergence. *a*, Sea floor tectonics off Central America. The East Pacific Rise (EPR) forms an impediment to subduction at the locus of ridge-trench collision (X), causing the Cocos plate (CO) to 'pivot' about a pole centred in the Gulf of California. Rotation is implied by the pattern of arcuate transforms and fanning of magnetic anomalies away from the pole of rotation. *b*, Sunda Arc (SA) plate tectonics. Oblique consumption of oceanic crust (Indian plate, IP) under the western Sunda Arc has led to the displacement of a major crustal wedge along a pair of transcurrent faults in the overriding plate. Displacement is in the direction of the transverse component of oblique convergence. SU, Sumatra. *c*, Rotation of Troodos oceanic crust. Troodos Ridge (TR) in Campanian located in 'pre-arc' or 'fore-arc' setting above northerly-dipping subduction zone (A). Ridge carried by subduction northward (B) against the Mamonia-Kyrenia margin (MK). Continued oblique consumption of crust to the south (A), consistent with the African-Eurasian plate convergence path³¹, imparts a torque on the overriding plate, driving the Troodos microplate anticlockwise, with the possible additive effect of pivoting at a ridge-transform-trench triple junction.

T.M.M.C. was supported by a NERC studentship. We thank the Cyprus Geological Survey, particularly C. Xenophontos and A. Panayiotou, for their assistance 'on the ground' in Cyprus.

Received 20 May; accepted 12 August 1985.

- Vine, F. J. & Moores, E. M. *Trans. Am. geophys. Un.* **50**, 131 (1969).
- Shelton, A. W. & Gass, I. G. *Ophiolites* (ed. Panayiotou, A.) (Printco, Nicosia, 1980).
- Moores, E. M. & Vine, F. J. *Phil. Trans. R. Soc. A* **225**, 417-467 (1971).
- Wilson, R. A. M. *Geol. Surv. Dep. Cyprus Mem.* **1** (1959).
- Kidd, R. G. W. & Cann, J. R. *Earth planet. Sci. Lett.* **24**, 151-155 (1974).
- Simonian, K. & Gass, I. G. *Bull. geol. Soc. Am.* **89**, 1220-1230 (1977).
- Verosub, K. L. & Moores, E. M. *J. geophys. Res.* **86**, 6335-6349 (1981).
- Blome, C. D. & Irwin, W. P. *Geology* **13**, 401-404 (1985).
- Mantis, M. *Bull. Sci. Res. Centre Cyprus, Epithris* **3**, 227-241 (1970).
- Robertson, A. H. F. & Hudson, J. D. in *Spec. Publ. int. Ass. Sedimentologists* Vol. 1 (eds Hsu, K. J. & Jenkyns, H. C.) 403-436 (Blackwell, 1974).
- Heller, F. J. *Geophys.* **42**, 475-488 (1977).
- Laj, C. *et al. Tectonophysics* **86**, 45-67 (1982).
- Lauer, J.-P. *Spec. Publ. geol. Soc. Lond.* **17**, 483-492 (1984).
- Van der Voo, R. & French, R. B. *Earth Sci. Rev.* **10**, 99-119 (1974).
- Robertson, A. H. F. *J. geol. Soc. Lond.* **134**, 269-292 (1977).
- Robertson, A. H. F. *J. geol. Soc. Lond.* **133**, 447-466 (1977).
- Ealey, P. J. & Knox, G. J. *Geologie Mijnb.* **54**, 85-100 (1975).
- Robertson, A. H. F. & Woodcock, N. H. *Bull. geol. Soc. Am.* **90**, 651-665 (1979).
- Robertson, A. H. F. & Woodcock, N. H. *Trans. R. Soc. Lond. Spec. Publ.* (in the press).
- Swarbrick, R. E. in *Ophiolites* (ed. Panayiotou, A.) 86-92 (Printco, Nicosia, 1980).
- Atwater, T. *Bull. geol. Soc. Am.* **81**, 3513-3526 (1970).
- Riddihough, R. P. *Can. J. Earth Sci.* **14**, 384-396 (1977).
- Menard, H. W. *J. geol.* **86**, 99-110 (1978).
- Fitch, T. J. *J. geophys. Res.* **77**, 4432-4461 (1972).
- Beck, M. E. *Tectonophysics* **93**, 1-11 (1983).
- Clube, T. M. M. & Robertson, A. H. F. *Surveys in Geophysics* (Reidel, Dordrecht, in the press).
- Schminke, H.-U. *et al. Geology* **11**, 405-409 (1983).
- Robinson, P. T. *et al. Geology* **11**, 400-404 (1983).
- Robertson, A. H. F. & Dixon, J. E. *Spec. Publ. geol. Soc. Lond.* **17**, 1-74 (1984).
- Pearce, J. A. *et al. Spec. Publ. geol. Soc. Lond.* **16**, 77-94 (1984).
- Livermore, R. A. & Smith, A. G. *Geology of the Taurus Belt* (eds Tekeli, O. & Gönçüoğlu, M. C.) 1-10 (MTA, Ankara, 1984).

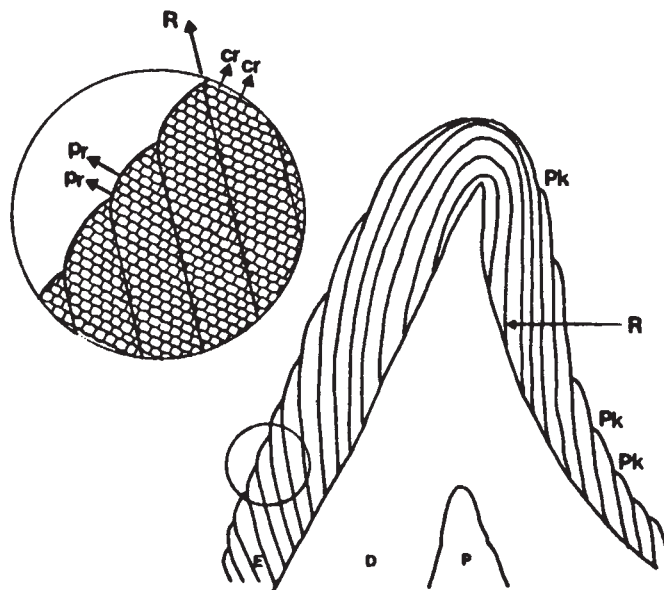


Fig. 1 Simplified diagram illustrating a bucco-lingual section through the enamel (E), dentine (D) and pulp (P) of an unworn incisor. The coarser incremental growth markings, striae of Retzius (R) can be seen passing obliquely from the enamel dentine junction to the tooth surface where they are visible as perikymata (Pk). Early growth increments around the incisal edge do not reach the enamel surface. The circular inset shows the finer enamel prisms (Pr) passing from the enamel-dentine junction across the striae of Retzius towards the enamel surface. On average seven prism cross-striations (Cr) (1 week's growth) are represented between adjacent striae of Retzius.

Re-evaluation of the age at death of immature fossil hominids

Timothy G. Bromage*† & M. Christopher Dean*

* Department of Anatomy and Embryology, University College London, London WC1E 6BT, UK

† Department of Anthropology, University of Toronto, Toronto, Canada M5S 1A1

We report here revised chronological ages at death of immature fossil hominids demonstrating for the first time that Plio-Pleistocene hominids had markedly abbreviated growth periods relative to modern man, similar to those of the modern great apes. Previous estimates of age at death for fossil hominids have principally been based on dental eruption, maturation and dental wear criteria for modern man^{1,2} and hence reflect their age in 'human' years³. We are now able to estimate the absolute duration of permanent incisor crown formation by observing gross incremental growth features in enamel and thereby apply a timescale to dental developmental events for specimens representing four Plio-Pleistocene fossil hominid taxa. Thus we have derived more reliable—species-specific—estimates of age at death that provide a more secure model on which to base studies of the palaeodemography, growth and maturation of early hominids.

Two types of incremental growth lines are present within enamel; enamel prism cross-striations and striae of Retzius (Fig. 1). Cross-striations occur along the length of the enamel prisms and have a periodicity of 4-8 μm in humans⁴. It is generally accepted that they result from a circadian variation in the rate of matrix secretion by ameloblasts²⁹. Boyde previously counted enamel cross-striations in ground sections of teeth and calculated the number of days between birth (neonatal line) and death of an individual in dentally immature human archaeological material^{4,5}. Striae of Retzius are coarser lines that pass obliquely from the enamel-dentine junction to the surface of the enamel where they become visible as perikymata^{6,7}, and are believed to be caused by an unidentified but regular and repeti-

tive systemic disturbance⁸. Newman and Poole⁸ report, from examinations of ground sections of human teeth, that although the amount of enamel between striae of Retzius varies, there are consistently 7-8 cross-striations between adjacent striae of Retzius which suggests that they are formed with a periodicity of 7-8 days. This close relationship between striae of Retzius and prism cross-striations is well noted in the literature^{9,10}. Asper⁶ demonstrated this relationship in the teeth of modern man, yet noted an occasional variation of 5-10 cross-striations between adjacent striae in a single tooth. There is further support for incremental periodicity in other hard tissues in man, reptiles¹¹ and birds¹². For example, in dentine, von Ebner's lines, Owens lines and peri-radicular bands, respectively, are equivalent to cross-striations, striae of Retzius and perikymata in enamel⁸. Bradford¹³ has also demonstrated a rhythmic disturbance in developing dentine indicating a 6-8-day repeat interval, and Yilmaz *et al.*¹⁴ have similarly confirmed a circadian rhythm in developing pig dentine by reporting 13 sets of alternating light and dark lines between markers of lead acetate injected at 2-week intervals. A regular 7-day rhythm has also been noted for growing human bone and confirmed experimentally in avian bone¹². Thus evidence for incremental periodicity generally exhibited in hard tissues gives credence to the observation of an average 7-8-day periodicity in enamel formation between any two adjacent perikymata.

Perikymata are often visible on human and hominid permanent incisors from the mammelons incisally to the enamel-dentine junction cervically. They cannot be seen on the surface of great ape teeth, but preliminary studies indicate that counts of striae of Retzius in great ape crowns provide estimates of crown formation times that are close to those predicted in the literature¹⁵. In other permanent teeth, perikymata do not reach the cuspal regions and are only visible over a reduced area of the enamel surface. For this reason permanent incisors were adopted for use in this study. In all teeth, however, some early growth increments of enamel are hidden beneath the surface and therefore do not become visible as perikymata. Examination of ground sections of unworn human incisors indicates that the

Table 1 Age reassessments based on hominid tooth samples

Specimen*	No. of perikyma†	Crown formation time‡	Root formation time	Revised age§	Dental formation (human years)
<i>Australopithecus afarensis</i> LH2 RI ₁	130	3 yr 0 mth	—	3.25	4.45
<i>Australopithecus africanus</i> Sts24a RI ¹	135	3 yr 1.1 mth	—	3.3	6.05
<i>Paranthropus robustus</i> SK62 RI ₁	57	1 yr 7.2 mth	1 yr 6 mth	3.35	5.85
LI ₂	64	1 yr 8.8 mth	1 yr 6 mth	3.48	
SK63 RI ₁	86	2 yr 1.8 mth	1 yr	3.15	6.10
SK71¶ RI ₂	61 #	1 yr 8.1 mth			
SK73¶ RI ₁	79 #	2 yr 0.2 mth			
Early <i>Homo</i> KNM-ER820 LI ₂	105	2 yr 6.2 mth	2 yr 6 mth	5.3	7.10
SK-74b† RI ₁	110	2 yr 7.4 mth			

* Specimens Sts24a (RI₁), Sts52 (L&RI₂), SK63 (RI₂), SK52 (RI₂), SK68 (LI₁) and SK69 (LI₁) were also investigated but did not show perikymata over the whole crown surface.

† In addition to the importance perikymata have in ageing studies, their spacing on the incisor crown surface appears to differentiate certain hominid taxa. Modern human teeth are reported as having perikymata that are more widely spaced toward the occlusal surface becoming more closely packed toward the cervix^{9,27}. This pattern also appears in the study sample of teeth of *Australopithecus* and early *Homo*. The teeth of *Paranthropus* observed in this study, however, do not show such a marked narrowing and condensation of perikymata cervically. An isolated lower incisor from Swartkrans, SK-74b, demonstrates the pattern seen in *Homo* and *Australopithecus* and contrasts with the pattern seen in *Paranthropus*. This evidence for a distinct patterning of perikymata suggests that SK-74b does not belong to *Paranthropus*²⁸ but belongs rather to early *Homo* (*Australopithecus* being unknown from Swartkrans).

‡ Crown formation times have been calculated using a 7-day cross-striation repeat interval between adjacent striae of Retzius and include 6 months growth prior to visible perikymata. Although variation in cross-striation repeat interval between adjacent striae of Retzius exists, an average 7-day incremental periodicity has been used, as this is most often cited in the literature.

§ Ages include the addition of 3 months for the time between birth and onset of incisor crown formation.

|| Kindly made available by Mark F. Skinner (unpublished manuscript).

¶ Isolated teeth.

Minimum counts that may be 1-3 less than the true figure due to abrasion at the cervix.

first 20 to 30 striae of Retzius are formed within the enamel before the remaining striae appear as perikymata. This indicates that enamel is forming for about 6 months before subsequent growth increments can be counted on the enamel surface. It is also well known that human central and lower lateral permanent incisors begin to calcify about 3-4 months after birth¹⁶ and that calcification of these teeth may occur even nearer to birth in the great apes¹⁵. However, estimates of chronological age at death of an individual must include the time of crown formation represented by the visible increments of growth (perikymata), estimates of early hidden growth increments and the short period between birth and the onset of calcification of the tooth. These estimates have been averaged at 3 months for hominoids between birth and onset of calcification and at 6 months for hidden growth increments in hominoid incisors. This gives a total of 9 months postnatal growth before the growth increments become visible at the incisal mammelons.

Some fossil hominid incisors in this study have some root formed and estimates for the time of root formation must also be included in order to calculate chronological age. There is evidence that Plio-Pleistocene hominids had a foreshortened period of root growth compared with modern man¹⁷. These estimates, therefore, have been based upon data on root growth in hominoids^{15,17,18}, with due regard to the relative rates of crown formation of the specimens and reference to radiographs^{19,20} of the specimens, in order not to impose human values for incisor root growth and development upon 'great ape-like' incisor crowns and developing dentitions.

The jaws of 33 immature fossil hominids from East and South Africa were replicated with a high-resolution silicone-based dental impression material and cast in epoxy resin²¹ as part of a facial remodelling and maturation study. Amongst the associated dental replicas, 11 incisors were selected as potentially useful for this study. Five additional incisors from the sample of isolated teeth from the Swartkrans hominid site were also selected for replication. Positive epoxy resin replicas were prepared for conventional scanning electron microscopy and

observed. Montage photographs were taken of the buccal surfaces of the teeth which were used to count the total number of perikymata on the tooth crown. Perikymata were visible from the unworn crown tip to the cervix in only nine specimens. The hominid sample is summarized in Table 1. Ten unworn lower incisor crowns of modern *Homo sapiens* were also prepared, observed and the perikymata counted. Estimates for the duration of crown formation in the modern human sample (mean number of perikymata 188 (range 165-202); mean crown formation time 4.2 yr) agree with estimates of crown formation times calculated from longitudinal and cross-sectional radiographic studies^{22,23}.

The assumption underlying the age reassessments presented in Table 1 is that they are chronological ages at death that have also been taken to represent the mean dental formation status for that age in each species. It is also assumed that because of the conservative process of dental development, any similarity in dental development between taxa mirrors similarity in growth and development²⁴. The ages derived for *Australopithecus*, *Paranthropus* and early *Homo* describe biological equivalence to modern man at roughly two-thirds the chronological age, demonstrating that they had growth periods similar to the modern great apes^{25,26}. Unexpectedly, the age calculated for early *Homo*, as represented by KNM-ER 820, provides evidence that the trend for prolongation of the growth period had not begun with this group. These results reflect a dramatic advance in our appreciation of the biology of these species which have, until now, been regarded in 'human' years.

Future studies should be aimed at confirming the periodicity of enamel incremental features in human and hominoid teeth. Unworn fully formed incisor crowns from humans of known age are largely unavailable (although tetracycline labelled teeth may prove useful) and perikymata on hominoid teeth are incompletely expressed, as noted above. However, it is still possible to examine cross-striation repeat intervals between striae of Retzius histologically, and this work is now in progress. Our finding, that the growth periods for Plio-Pleistocene hominids were similar to the modern great apes implies that little

taxonomic significance can be attributed to this aspect of the biology of early hominids during these phases of human evolution. However, future estimates of age at death for the majority of sub-adult early Plio-Pleistocene fossil hominids, based on this study and new dental developmental data, should help quantify the nature of changes in rates of growth in specific anatomical regions, and enable these changes to be traced through the hominid fossil record.

We are particularly grateful to Professor Alan Boyde, of The Hard Tissue Unit, University College London, for both his helpful comments and for providing facilities to carry out this research. We also thank P. Andrews, A. D. Beynon, E. Delson and S. J. Jones for their comments on the manuscript, R. H. Herold, W. H. Kimbel, L. B. Martin, S. A. Reid, M. F. Skinner, C. B. Stringer, T. Watson and B. A. Wood for their helpful comments and the Governments of Kenya and Tanzania, C. K. Brain, A. R. Hughes, M. D. Leakey, R. E. F. Leakey, P. V. Tobias and E. Vrba for permission to study hominid material in their care. Support for this work was provided by the L.S.B. Leakey Foundation and University of Toronto Doctoral grants (T.G.B.) and the Wellcome Trust and Nuffield Foundation (M.C.D.).

Received 7 March; accepted 29 August 1985.

1. Mann, A. E. *Some Paleodemographic Aspects of the South African Australopithecines* (University of Pennsylvania, Philadelphia, 1975).

2. Skinner, M. F. & Sperber, G. H. *Atlas of Radiographs of Early Man* (Liss, New York, 1982).
3. Skinner, M. F. thesis, Univ. Cambridge (1975).
4. Boyde, A. *Excerpta medica int. Congr. Ser.* **80**, 36–46 (1963).
5. Miles, A. E. W. in *Dental Anthropology* (ed. Brothwell, D.) 191–202 (Pergamon, Oxford, 1963).
6. Asper, V. H. *Schweiz. Vjschr. Zahnheilk.* **26**, 275–314 (1916).
7. Risnes, S. *Scand. J. dent. Res.* **92**, 1–5 (1984).
8. Newman, H. N. & Poole, D. F. G. *Archs oral Biol.* **19**, 1135–1143 (1974).
9. Gustafson, G. & Gustafson, A.-G. in *Structural and Chemical Organization of Teeth II* (ed. Miles, A. E. W.) 75–134 (Academic, London, 1967).
10. Poole, D. F. G. in *Structural and Chemical Organization of Teeth I* (ed. Miles, A. E. W.) 111–149 (Academic, London, 1967).
11. Poole, D. F. G. *Q. Jl microsc. Sci.* **98**, 349–367 (1957).
12. Sissons, H. A. *Nature* **163**, 1001–1002 (1949).
13. Bradford, E. W. in *Structural and Chemical Organization of Teeth II* (ed. Miles, A. E. W.) 3–34 (Academic, London, 1967).
14. Yilmaz, S., Newman, H. N. & Poole, D. F. G. *Archs. oral Biol.* **22**, 511–513 (1977).
15. Dean, M. C. & Wood, B. A. *Folia primatol.* **36**, 111–127 (1981).
16. Logan, W. H. G. & Kronfeld, R. J. *Am. dent. Ass.* **20**, 379–427 (1933).
17. Dean, M. C. *Am. J. phys. Anthropol.* (in the press).
18. Moorrees, C. F. A., Fanning, E. A. & Hunt, E. E. Jr *J. dent. Res.* **42**, 1490–1502 (1963).
19. Abbott, S. A. thesis, Univ. London (1984).
20. Leakey, R. E. F. & Wood, B. A. *Am. J. phys. Anthropol.* **39**, 355–368 (1973).
21. Bromage, T. G. *J. Microsc.* **137**, 209–216 (1985).
22. Nolla, C. M. *J. Dent. Child.* **27**, 254–266 (1960).
23. Fass, E. N. *J. Dent. Child.* **36**, 391–401 (1969).
24. Swindler, D. R. in *Nonhuman Primate Models for Human Growth and Development*, 67–94 (Liss, New York, 1985).
25. Gavan, J. A. & Swindler, D. R. *Am. J. phys. Anthropol.* **24**, 181–190 (1966).
26. Bromage, T. G. in *The Past, Present, and Future of Hominoid Evolution* (ed. Tobias, P. V.) (Liss, New York, in the press).
27. Boyde, A. in *The Prevention of Periodontal Disease* (eds Eastoe, J. E., Picton, D. C. A. & Alexander, A. G.) 46–63 (Kimpton, London, 1971).
28. Robinson, J. T. *Trans. Mus. Mem.* **9** (1956).
29. Boyde, A. in *Scientific Foundations of Dentistry* (eds Cohen, B. & Kramer, R. A.) 335–353 (1976).

Subjective contours capture stereopsis

V. S. Ramachandran* & P. Cavanagh†

* Department of Psychology, University of California, San Diego, La Jolla, California 92093, USA

† Departement de Psychologie, Université de Montréal, Montreal, Quebec, Canada H3C 3J7

Stereoscopic depth perception is based on measuring tiny differences between the two eyes' images which arise as a result of binocular parallax¹. Julesz² used random-dot stereograms to demonstrate that stereopsis may be based on a simple point-to-point comparison of the two eyes' images and does not require the presence of monocularly visible forms or contours. Here we present a new class of stereograms which illustrate that monocular subjective contours^{3–5} can influence the matching of elements in a stereogram even though the elements themselves convey no disparity information. More specifically, the depth seen from such contours is automatically attributed to texture elements and lines that are enclosed by these contours—an illusion that we call 'stereoscopic capture'.

Figure 1a depicts a well-known illusion known as the 'wallpaper effect'. When patterns consisting of repeating stripes are viewed binocularly they can convey any one of a number of different depth planes⁶. However, at any given instant the entire pattern is seen to occupy only one single plane; the exact plane seen seems to depend mainly on the angle of convergence. Figure 1b shows a square defined by subjective contours^{3–5}. Such contours can be produced by means of appropriately aligned black disks from which right-angled sectors have been removed. The brain interprets this figure parsimoniously as an opaque white square with its four corners occluding the four black disks (and not as four sectorized disks which have been deceitfully aligned by the experimenter). One has the strange impression of a contour connecting these aligned edges even though no contour exists physically—hence the name subjective contours. Whether these contours are physical, physiological or truly 'subjective' is a much debated semantic issue that need not concern us here.

We constructed a stereogram using two subjective squares similar to that of Fig. 1b by introducing small horizontal disparities between the vertical edges of the cut-out sectors (Fig.

1c). Perception of depth in these stereograms is most vivid when the disparities convey a square standing out in front of the disks. If the two images are now interchanged, stereopsis disappears and, surprisingly, the subjective contours also become weak; perhaps because it is now difficult to make any sense out of the whole figure^{7,8}. These stimuli and all subsequent ones described here were generated on a CRT screen using a Macintosh microcomputer. The visual angle subtended by the black disks was 0.5° and the subjective square itself subtended 2°. The horizontal disparity between the vertical edges of the cut sectors was 20 arc min.

Our next step was to superimpose a template of Fig. 1c on several kinds of wallpaper to generate 'wallpaper stereograms' (Figs 2, 3, 4). In Fig. 2 we simply used repeating vertical rows of spots (see figure legends for additional details). This stereogram was then shown to eight subjects who were experienced psychophysical observers but were unaware of the purpose of the experiment. They were shown both crossed and uncrossed versions (that is, with a square depicted in front or behind the plane of fixation) and asked to report what they saw. With crossed stereograms (Fig. 2a) all eight subjects saw the square defined by the subjective contours as standing out clearly in front. Interestingly, the corresponding dot-rows on the wallpaper were also carried forward with the plane—an effect we call 'stereoscopic capture'. As the disparity of the squares equalled several multiples of the periodicity of the dot-rows, this finding implies that the subjective surface in depth created by the subjective contours was somehow pulling the dots with it even though the dots themselves do not convey any specific disparity information (in fact they convey zero disparity in relation to the background). When the subjects viewed uncrossed stereograms, no such effect was seen; neither the subjective surface nor the stereoscopic capture effect was observed (Fig. 2b; see legend for further details).

In Fig. 3 we used continuous vertical lines instead of rows of dots in the background. Subjects reported that the illusion was just as compelling here as in Fig. 2. Surprisingly, the capture effect was strong enough to overcome the physical continuity of the vertical lines and caused apparent breaks to appear on the lines at the upper and lower edges of the subjective square. Again, as in Fig. 2, the illusion disappeared completely when the pictures received by the two eyes were interchanged and subjects now reported seeing rivalry and diplopia instead of capture.

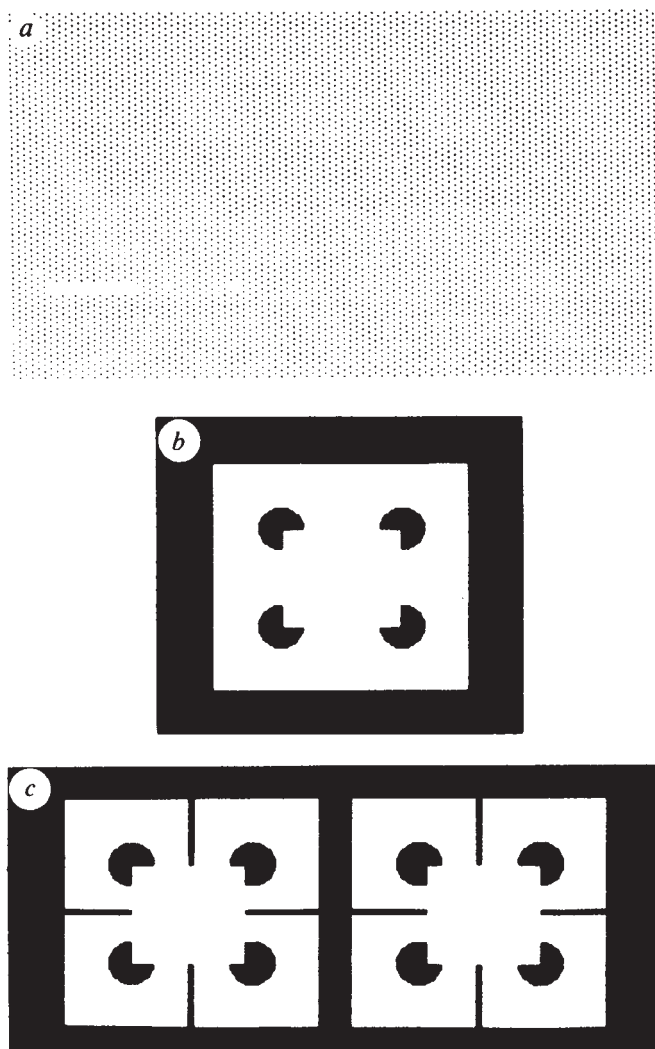


Fig. 1 *a*, A well-known stereoscopic illusion called the wallpaper effect. If the reader brings the pattern very close to his nose and changes his angle of vergence while viewing this display he will perceive corresponding changes in the plane of perceived stereoscopic depth. *b*, A square defined by subjective contours. Observers usually perceive an opaque white square whose corners partially occlude the four black disks. Faint 'subjective' lines are seen to delineate the square even though no physical lines exist. *c*, A stereogram produced by using two subjective squares similar to that in *b*. Small horizontal disparities are introduced between the vertical edges of the cut sectors. When the two eyes' patterns are fused, a subjective square stands out in stereoscopic depth.

Next we superimposed Fig. 1c on horizontal lines (Fig. 4a). As expected, none of the eight subjects observed capture when uncrossed disparities were used (as in Fig. 3). With crossed disparities, subjects reported that only the lines actually adjoining the sectors themselves were pulled forward. All the other lines on the square remained flush with the background and it was very difficult to break the continuity of these lines in order to partition them into two surfaces. Perhaps the brain is reluctant to attribute depth to horizontal lines as such lines do not normally convey disparity signals in the real world.

One could argue that the anisotropy observed here is between vertical as opposed to horizontal gradients of disparity (that is, between the vertical and horizontal edges of the illusory square) and not between vertical and horizontal gratings. We tested this possibility by using oblique gratings (Fig. 4b) and found that the illusory breaks on the lines could now be seen along both vertical and horizontal borders of the square. Further, when they were asked to compare the vertical and horizontal depth edges, none of our subjects could see any difference. This result

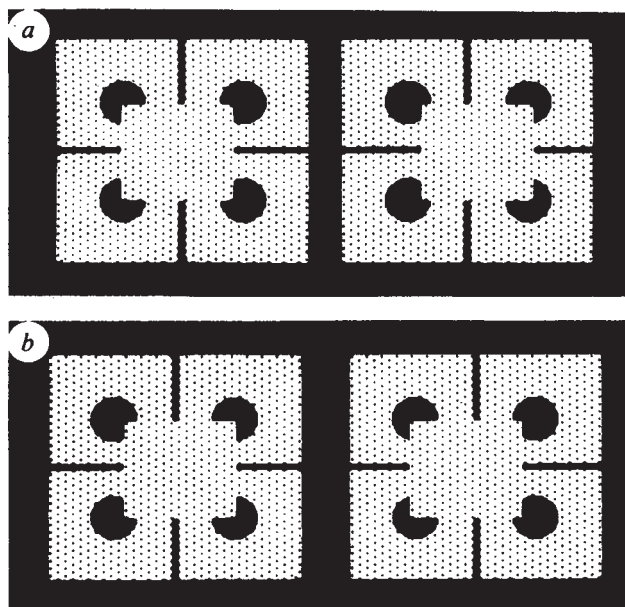


Fig. 2 *a*, Produced by superimposing a template of Fig. 1c on a repeating wallpaper pattern. The inter-dot separation between the elements constituting the wallpaper is 5 arc min. A subjective square stands out in front of the background and carries the wallpaper with it even though the elements on the wallpaper are at zero disparity. This is an example of stereo capture (see text). *b*, Identical to *a* but with the two eyes' pictures interchanged to reverse retinal disparities. The illusion is destroyed and replaced by rivalry and diplopia. However, one of our subjects spontaneously reported that instead of stereo capture he now saw four portholes cut out of an opaque sheet of wallpaper. Through these portholes he could see the four corners of a smaller square piece of wallpaper and these corners pulled the corresponding rows of dots with them. This alternative and unusual percept of seeing four deeper corners is of considerable interest for it reinforces the view that it is indeed the subjective contours that drive the perceived depth. The observation was also confirmed by two other subjects on prompting but the remaining five subjects continued to experience rivalry and diplopia. The percept becomes more pronounced if the horizontal black lines are removed and if finer textures are used (for example, try fusing Fig. 3b).

suggests that the anisotropy observed in Figs 3 and 4a occurs not at the cyclopean level but probably at an earlier stage of stereo-matching.

Figure 4c depicts another version of stereo capture: here we have used dissimilar wallpaper textures for the images received by the two eyes. As the elements comprising these textures are uncorrelated for the two eyes, one would expect to see either binocular rivalry or random depth planes based on chance correlations. Seven of our eight subjects, however, reported seeing stereo capture; the dots corresponding to the square were seen to stand out in front despite the rivalry. Thus, the stereo capture illusion is strong enough to overcome rivalry of finer image features (high spatial frequencies), although the effect is certainly less striking than that observed in Figs 2, 3. This observation is consistent with the conclusions of Ramachandran *et al.*⁹ that monocularly visible texture boundaries can support stereopsis even when the elements defining the textures are themselves uncorrelated.

We considered the possibility that vergence eye movements are involved in producing these effects—this seems unlikely for two reasons. First, one sees two different stereoscopic planes simultaneously in each of these stereograms, an outcome that would be impossible to achieve by simply using convergence alone. Second, we presented the stereograms (Figs 2, 3b) in a tachistoscope to three subjects. Crossed and uncrossed disparities were flashed in random sequence while they fixated a spot presented on the background just below the subjective square. When asked to report whether the elements within the

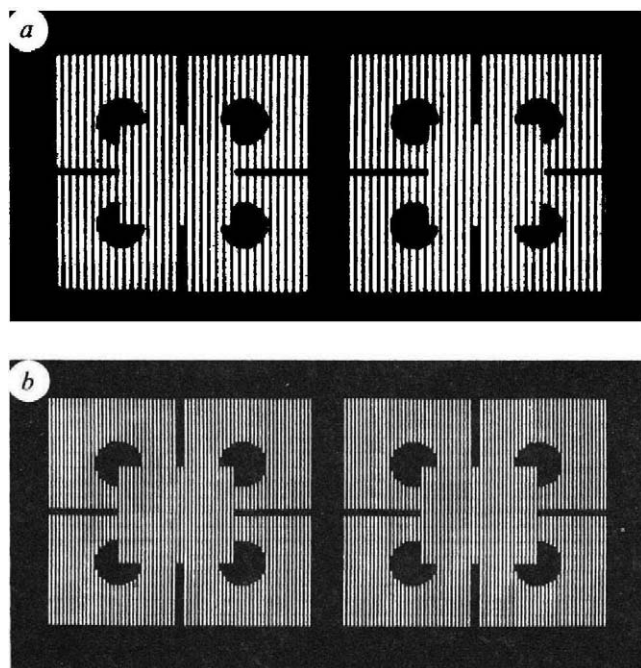


Fig. 3 *a*, A template of Fig. 1c superimposed on a pattern of repeating vertical stripes (spatial frequency of the stripes is 9 cycles deg^{-1}). A subjective square stands out clearly from the background, carrying the lines with it. Also, the capture effect is strong enough to overcome the physical continuity of the lines and causes apparent breaks to appear on the lines. *b*, Similar to *a* except that the disparity has been reduced to 10 arc min to enable easier fusion. The spatial frequency of the grating is 18 cycles deg^{-1} .

square appeared in front of or in the same plane as the fixation spot, they responded correctly on 51 out of 60 trials (3 subjects $\times$ 20 trials each). Similar results were obtained with Fig. 3b (54 out of 60 responses were correct), suggesting that eye movements are not responsible for the effect.

We found that stereoscopic capture was also very sensitive to the spatial phase relationship between the wallpaper and the sectorized disks: the cut edges had to be 'phase-locked' with the lines so that the subjective square's disparity was an exact multiple of the periodicity of the lines. Thus, when the disks themselves were flush with the background, the subjective square occupied exactly the same plane as the enclosed wallpaper that was captured and this seemed to amplify the illusion. If a deliberate phase shift was introduced, capture was considerably reduced and replaced by the appearance of a transparent subjective square. This observation implies that the signal derived from the disparate subjective contours is not merely attributed to the elements of wallpaper; there must be some degree of mutual synergy between the two. In this respect stereo capture may turn out to be different from motion capture^{10,11}.

Our conclusions are also consistent with the elegant demonstrations of Mitchison and McKee¹² and Julesz and Chang¹³, although neither of these studies specifically explored the role of image segmentation produced by illusory contours. Mitchison and McKee found that disparity signals derived from the endpoints of two horizontal dot-rows viewed binocularly could influence the matching of ambiguous repetitive elements in the middle. However, in certain conditions they could produce a dot-row that was tilted in stereoscopic space as a result of interpolation—an effect that we could not produce in our displays. When we superimposed a tilted illusory surface on closely spaced gratings (12 cycles deg^{-1}) the gratings remained flush with the background and showed no tendency to tilt with the illusory surface.

Next, we wondered whether the stereo capture illusion depicted in Figs 2 and 3 was being produced by the illusory contours themselves or by the disparity conveyed by the cut sectors. Although there is no simple way of answering this question, we

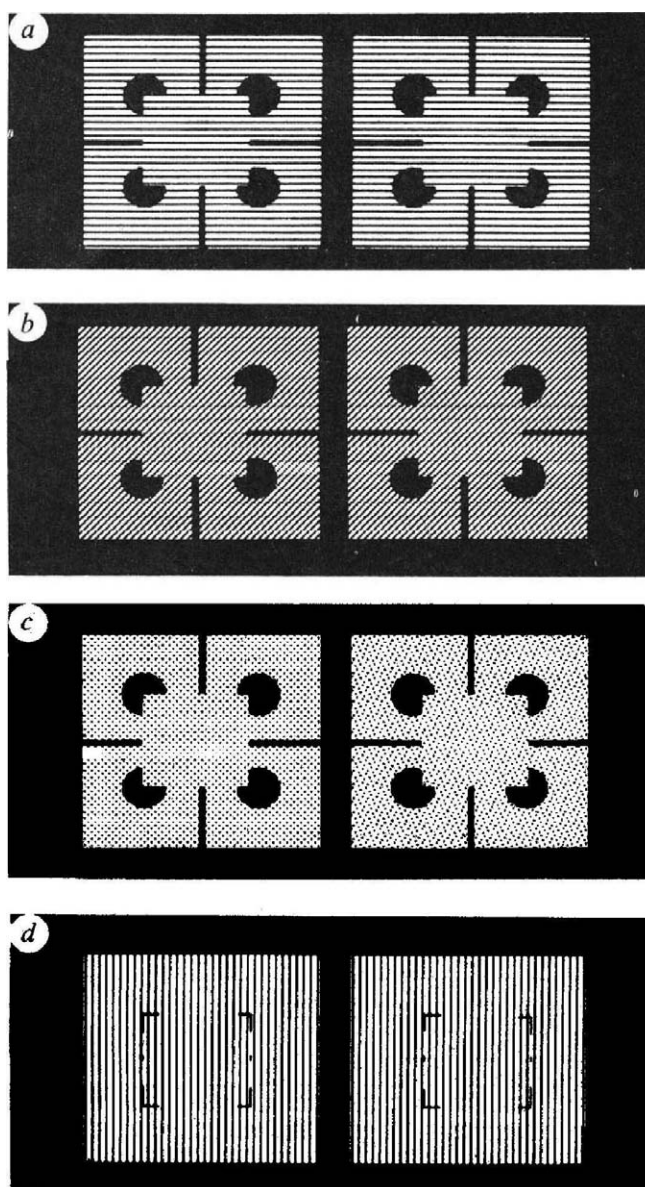


Fig. 4 *a*, Capture is considerably weaker if horizontal lines are used. The lines adjoining the sectors themselves are pushed forward but not the lines in between (they remain flush with the background and it is impossible to see subjective breaks). *b*, A stereogram similar to *a* and Fig. 2, except that we have used oblique lines instead of horizontal or vertical lines. The apparent breaks on the lines are equally obvious on both vertical and horizontal borders of the illusory square. *c*, Here we have used dissimilar textures for the two eyes to produce rivalry. Stereo capture can be seen despite the rivalry (especially on eccentric fixation) but the illusion is less compelling than in Figs 2 and 3. Note that although the textures are dissimilar, they probably have some overlap of their Fourier power spectra; this overlap may produce some chance correlations that are necessary for generating the effect, but we did not specifically explore this possibility. *d*, This stereogram is identical to Fig. 3a in every respect except that we have replaced the cut sectors with corners made of small line segments so that no illusory contours are visible. Although the disparity between these corners is identical to the disparity between the cut sectors in Fig. 3a, stereoscopic capture cannot be observed in this display, suggesting that the presence of an illusory surface is an important prerequisite. Some of the lines may occasionally appear to come forward due to convergence but it is impossible to partition the image into two well-defined surfaces.

attempted to do so by using a 'control' stereogram of the kind shown in Fig. 4d. Here, we have replaced the cut sectors with corners made of tiny line segments so that no illusory contours are visible. When our eight subjects viewed this stereogram, they

reported that the corners floated out in front of the paper but that the grating was not captured and pulled forward. This supports our contention that the creation of a subjective surface is an important prerequisite for producing stereo capture.

Thus, disparate subjective contours capture regions of zero disparity enclosed within these contours but not regions that lie outside the contours. When vertical stripes are used the effect is strong enough to overcome the physical continuity of the lines so that illusory breaks appear on them. The capture effect disappears if the sign of disparity is reversed, suggesting that factors such as overlay or occlusion are an integral part of the illusion. An anisotropy was observed between vertical and horizontal lines, the latter being resistant to capture. Capture was also phase-sensitive, the effect being most pronounced when the disparity of the illusory contours was an exact multiple of the grating periodicity. Finally, completely uncorrelated (and rivalrous) regions can also be captured to a limited extent.

It would be interesting to determine whether computational models¹⁴ of stereopsis can be modified to account for these intriguing phenomena. We conclude that in many instances the brain may begin by segmenting the scene into contours and surfaces and that the information derived from such segmentation can have a profound influence on subsequent processing retinal disparities⁹. Certain cells in area V1¹⁵⁻¹⁷ are thought to respond to small disparities and those in V2 to large retinal disparities¹⁸⁻²⁰ as well as illusory contours²¹. It is not inconceivable, therefore, that stereo capture results directly from synergistic interactions between these cells. For example, the wallpaper might excite cells corresponding to several depth planes in area V1 whereas the disparity of illusory contours would excite only a single plane in V2. The latter signal might then be fed back to V1 to select or 'highlight' the appropriate plane—a conjecture

that is consistent with the observed anisotropy (Figs 3, 4a) and phase-sensitivity of stereo capture. Further experiments along these lines may help us to better understand the hierarchy of precedence rules in the brain that result in the construction of a three-dimensional visual world.

We thank D. MacLeod, F. H. C. Crick and G. Mitchison for stimulating discussions, E. Ebbesen for facilities, and one of the referees for suggesting some further experiments. V.S.R. was supported by academic senate and biomedical research grants from UCSD.

Note added in proof: If disparities were introduced between the disks themselves and not just between the cut sectors, no capture was seen. This suggests that a simple spread of disparity signals cannot explain the effect; the presence of an illusory stereoscopic surface is required.

Received 11 April; accepted 18 July 1985.

1. Wheatstone, C. H. *Phil. Trans. R. Soc.* **128**, 371-394 (1833).
2. Julesz, B. *Foundations of Cyclopean Perception* (University of Chicago Press, 1971).
3. Schumann, F. Z. *Psychol.* **1**, 23, 1-32 (1904).
4. Kanizsa, G. *Riv. Psicol. norm. patol. appl.* **49**, 7-30 (1955).
5. Rock, I. *The Logic of Perception* (MIT Press, London, 1983).
6. Lie, I. *Scand. J. Psychol.* **6**, 109-116 (1965).
7. Blomfield, S. *Nature new Biol.* **245**, 147 (1973).
8. Harris, J. P. & Gregory, R. L. *Perception* **2** (2), 235-247 (1973).
9. Ramachandran, V. S., Rao, V. M. & Vidyasagar, T. R. *Nature* **242**, 412-414 (1973).
10. Ramachandran, V. S. & Inada, V. *Spatial Vision* Vol. 1 (in the press).
11. Ramachandran, V. S. & Cavanagh, P. *Vision Res.* (submitted).
12. Mitchison, G. & McKee, S. *Nature* **315**, 402-404 (1985).
13. Julesz, B. & Chang, J. J. *Biol. Cybernet.* **22**, 107-119 (1976).
14. Marr, D. & Poggio, T. *Proc. R. Soc. B24*, 301-328 (1979).
15. Barlow, H. B., Blakemore, C. & Pettigrew, J. D. *J. Physiol., Lond.* **193**, 327-342 (1967).
16. Pettigrew, J. D., Nikara, T. & Bishop, P. O. *Expl Brain Res.* **6**, 391-410 (1968).
17. Fischer, F. & Poggio, G. F. *Proc. R. Soc. B204*, 409-414 (1979).
18. Hubel, D. H. & Wiesel, T. N. *Nature* **225**, 41-42 (1970).
19. Clarke, P. G. H., Donaldson, I. M. L. & Whitteridge, D. J. *J. Physiol., Lond.* **256**, 509-526 (1976).
20. Ramachandran, V. S., Clarke, P. G. H. & Whitteridge, D. *Nature* **268**, 333-335 (1977).
21. Van der Heydt, R., Peter Hans, E. & Baumgartner, G. *Science* **224**, 1260-1262 (1984).

Cycles of presence and absence of mother mouse entrain the circadian clock of pups

N. Viswanathan & M. K. Chandrashekar

Department of Animal Behaviour, School of Biological Sciences, Madurai Kamaraj University, Madurai—625 021, India

Recent findings on neural and endocrine rhythms in infant mice and rats show that maternal coordination has an important role in setting the phase of the developing circadian clock both in the fetus and soon after birth¹⁻⁷. However, less information is available about the influence of the mother on activity/rest cycles of infants. Separation of the mother from infants in guinea pigs, monkeys and rats⁸⁻¹⁰ results in an increase in sleep disturbance (enhanced activity?). In this context it may be a common feature that during the postnatal period there is enhanced activity of pups during the hours when the mother is not nearby. Conversely, the social influences exerted by the mother while present with her young possibly leads to a relative rest stage. We have now tested this assumption in the night-active mouse *Mus booduga*. Our study addressed the postulate that the circadian activity/rest cycles of the pups are controlled by cyclic(?) presence and absence of the mother. The results reported here clearly indicate that the circadian locomotor activity of pups kept under continuous illumination or continuous darkness do entrain them to a regime of imposed 12:12-h cyclic presence and absence of the mother. The characteristics of this entrainment confer on the mother mouse the role of zeitgeber.

Pregnant females of the field mouse *M. booduga* (14-16 g) were captured from the fields surrounding the university campus. The pregnant mice were maintained in continuous darkness (DD) at a temperature of $28 \pm 1^\circ\text{C}$ and littered two to eight pups. For the first series of experiments two pups of either sex were selected from each litter. The animals for these experiments

were from a total of 14 mothers and 28 pups. One pup was placed in continuous illumination (LL) of 0.5-0.7 lx incandescent light from day 5 of postnatal life whereas the other pup remained in DD. (A 12-h separation of the pups from the mother before day 5 resulted in excessive mortality.) Starting on day 5 until day 15, the mothers were presented for 12-h periods alternately to pups in LL during 06:00-18:00 h and to pups in DD during 18:00-06:00 h. The circadian rhythm of the mother mice would have been entrained to the imposed light/darkness (LD) cycles while being alternately berthed with the pups. Food and water were available at all times. By day 16 of postnatal life, pups of *M. booduga* can already run the activity wheel independently. Their locomotor activity was measured from day 16 onwards for about 4 weeks, individually using running wheels in the absence of the mother in the same lighting conditions as had prevailed before. The locomotor activity of the mother was also similarly recorded in DD from day 16 onwards. The revolutions of the wheels were monitored with an A620X Esterline Angus event recorder.

The mother mice began to be active at 18:00 h on day 1 of the recording (day 16 after littering the pups) and the activity rhythm free-ran thereafter on subsequent days. The circadian rhythms in the locomotor activity of the pups in LL and those in DD revealed onsets of activity on day 1 of the recording which were 180° off course relative to each other and free-ran on the subsequent days. For each pup, this onset of activity coincided with the time at which the mother had been removed during entrainment. Figure 1a-c illustrates these trends. From these results we infer maternal entrainment of the circadian clock of pups.

In a second series of experiments mother mice were again cyclically presented from day 5, but in this series the mice were presented beyond day 16 for a further 2-5 weeks. Sixteen days after littering, the mother mice were tethered by a 10-cm aluminium chain. The presence/absence (PA) cycles of the mother were continued as in the first series of experiments. The restricting chain meant that the mother mouse could not enter the activity wheel from the nesting cage whereas the pups could.

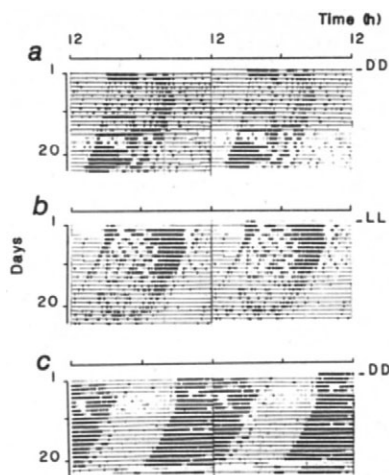


Fig. 1 Double-plotted wheel-running activity of a field mouse mother in DD (a) and two of her pups, one in LL (b) and the other in DD (c). Activity bouts deflected the writing stylets, causing bands on the strip charts, and rest traced mere lines. Day 1 of the records is identical with day 16 of the animals after parturition. Free-runs of the activity rhythms of the LL pup (b) and DD pup (c) start ~ 12 h (180°) apart.

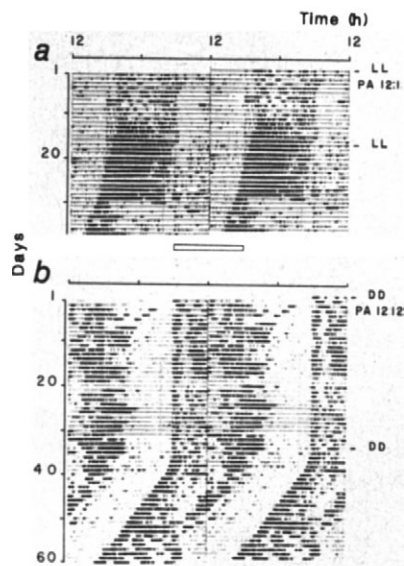


Fig. 2 Double-plotted wheel-running activity of the pups: a, days 1-18, LL and PA cycles of mother 12:12 h; days 19-37, LL. b, Days 1-35, DD and PA cycles of mother 12:12 h; days 36-60, DD. Presence of mother: LL pup, 06:00-18:00 h; DD pup 18:00-06:00 h. The presence of the mother is indicated by an open bar. Day 1 of the records is identical with day 16 of postnatal life.

It was thus possible to measure the locomotor activity of the pups alone. Four mothers and eight pups were used in these experiments.

Figure 2 illustrates the type of entrainment of the pups in LL and the pups in DD to the imposed 12:12-h PA cycles of the mother. During the days of entrainment, LL and DD pups ran the activity wheel only in the absence of the mother. Thus, the LL pups ran when the DD pups were huddled with the mother and vice versa. Consequently, entrainment in LL and DD pups ensues with the appropriate 180° displacement. When PA cycles of the mother were discontinued, the rhythm of the pups started to free-run with the onset of activity again coinciding with the time of removal of the mother during entrainment. The successful synchronization of the locomotor activity of the pups during PA cycles confirm the role of mother as zeitgeber. We propose that the social contact between mother and pup is taken by the

pup for subjective day (rest hours) and the absence of the mother is taken for subjective night (activity hours).

In natural conditions, the pups are probably never directly exposed to environmental LD conditions, as these mice are burrow-dwelling animals. The mother mice stay inside the burrow during the day and forage during the night. In the absence of any extra-retinal photoreceptors in mammals¹¹, perception of whatever levels of ambient light intensities are available in burrows can only be through the eyes. These pups, being altricial, do not open their eyes until 12-14 days after parturition⁷ and may thus be unable to perceive LD conditions even if there is light in the burrows. Under these circumstances, the pups may well rely behaviourally on the mother and take her presence for day and absence for night. Such mother/infant interactions probably help the pups to maintain the phase relationship of the prenatally set clock to that of the environment. This type of maternal contact is of ecological importance and helps to optimize the economy of the biological system and prepares the developing mammal for entry into the external environment.

We thank Dr R. Subbaraj for discussions and the DST, Government of India, for financial assistance. N.V. is a Senior Research Fellow of the CSIR.

Received 9 April; accepted 7 August 1985.

1. Deguchi, T. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2814-2818 (1975).
2. Fuchs, J. L. & Moore, R. Y. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1204-1208 (1980).
3. Hiroshige, T., Honma, K. & Watanabe, K. *J. Physiol., Lond.* **325**, 521-532 (1982).
4. Honma, S., Honma, K., Shirakawa, T. & Hiroshige, T. *Endocrinology* **114**, 1791-1796 (1984).
5. Reppert, S. M. & Schwartz, W. J. *Science* **220**, 969-971 (1983).
6. Takahashi, K., Murakami, N., Hayafuji, C. & Sasaki, Y. *Am. J. Physiol.* **246**, R359-R363 (1984).
7. Viswanathan, N. & Chandrashekar, M. K. *Proc. Indian Acad. Sci. (Anim. Sci.)* **B93**, 235-241 (1984).
8. Astic, L. & Jouvett-Mounier, D. *J. Physiol., Paris* **60**, 389 (1968).
9. Hofer, M. A. & Shair, H. *Dev. Psychobiol.* **15**, 229-243 (1982).
10. Reite, M. & Short, R. A. *Archs gen. Psychiat.* **35**, 1247-1253 (1978).
11. Rusak, B. & Boulos, Z. *Photochem. Photobiol.* **34**, 267-273 (1981).

Presynaptic serotonin receptor-mediated response in mice attenuated by antidepressant drugs and electroconvulsive shock

Guy M. Goodwin, Rosemarie J. De Souza
& A. Richard Green

MRC Clinical Pharmacology Unit, Radcliffe Infirmary,
Oxford OX2 6HE, UK

Ligand-binding studies have demonstrated two types of serotonin (5-HT) receptor, 5-HT₁ and 5-HT₂, in the brains of rodents¹ and there is additional evidence for the existence of 5-HT₁ subtypes^{2,3}. Recently a new drug, 8-hydroxy-2-(di-N-propylamino)tetralin (8-OH-DPAT), has been identified which shows high selectivity for binding to 5-HT₁ (possibly 5-HT_{1A}) receptors³ and which binds to presynaptic serotonin autoreceptors in some regions of rat brain^{4,5}. We have shown previously, that this compound produces a hypothermic response in mice^{6,7}, probably via an agonist action at serotonin presynaptic receptors. Here we show that a wide range of antidepressant treatments decrease the hypothermic response to 8-OH-DPAT over a time course comparable to the onset of therapeutic action. Interestingly, repeated electroconvulsive shock (ECS) has the same effect. We propose that this change is relevant to the mechanism of action of antidepressant drugs.

Male C57/B16/O1a mice (Olac, Bicester) were used throughout. The mice were housed in groups of eight in conditions of controlled temperature (20°C) and lighting (dark period: 1900-0700 h) and fed an *ad libitum* diet of 41B pellets and tap water.

The hypothermic response of the mice was determined by measurement of rectal temperature, all measurements being

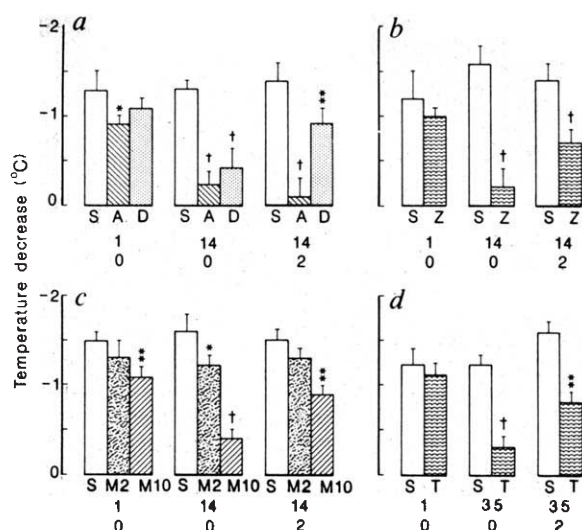


Fig. 1 Effect of antidepressant drug administration on the hypothermic response of mice to 8-OH-DPAT. *a*, Tricyclic antidepressants; *b*, zimeldine; *c*, mianserin; *d*, tranlycypromine. Mice were injected with amitriptyline (A, 20 mg kg⁻¹ i.p.), desipramine (D, 20 mg kg⁻¹ i.p.), zimeldine (Z, 20 mg kg⁻¹ i.p.), mianserin at either 2 mg/kg⁻¹ i.p. (M2) or 10 mg kg⁻¹ i.p. (M10), tranlycypromine (T, 6 mg kg⁻¹ i.p.) with control mice receiving saline i.p. (S). The mean temperature decrease $\pm$ s.e.m. (bars) 20 min following 8-OH-DPAT (0.5 mg kg⁻¹ s.c.) is shown. Data presented show results 24 h after 1 dose, 14 once-daily doses (or 35 doses in the case of tranlycypromine) and after 2 days withdrawal from chronic administration. Hypothermic responses are different from appropriate control as shown: **P* < 0.05; †*P* < 0.005; ‡*P* < 0.001. There are no differences in basal temperature between the control groups and drug-treated groups at any of the times examined.

made in a room with ambient temperature of 20 ± 0.5 °C. Temperature was measured just before injection of 8-OH-DPAT (0.5 mg per kg subcutaneously, s.c.), a dose that produces about half the maximum response⁷, and 20 min later, the nadir of the response at this dose⁷.

Groups of mice were injected intraperitoneally (i.p.) with antidepressant drugs and the temperature change following 8-OH-DPAT examined 24 h later. The mice were then given daily injections of antidepressant drugs for 13 more days; the effect of 8-OH-DPAT on rectal temperature was tested both the day after the last dose of antidepressant and after a further 48 h to examine the hypothermic response after withdrawal of the drug.

One day of amitriptyline administration produces a slight attenuation of the 8-OH-DPAT-induced hypothermia (Fig. 1*a*), although the attenuation following desipramine and zimeldine was not statistically significant (Fig. 1*a,b*). However, after 14 days of drug administration there is pronounced attenuation of the response following injection of all three drugs (Fig. 1). Indeed, the temperature decrease induced by 8-OH-DPAT is almost totally abolished both 20 min (Fig. 1) and 40 min. (data not shown) after 8-OH-DPAT. This attenuation is still present 48 h after drug withdrawal, the decreased response being at least as marked in the amitriptyline-treated mice (Fig. 1). Thus, drugs displaying selectivity in inhibiting serotonin uptake (zimeldine)⁸, noradrenaline uptake (desipramine)⁹ and having an almost equal potency in inhibiting both amine-uptake pumps (amitriptyline)¹⁰ all have the same effect. Therefore, we next examined mianserin, a drug that has little effect on amine uptake¹⁰. At a dose (2 mg per kg) that profoundly inhibits 5-HT₂ binding and head-twitch behaviour¹¹ it produces a modest attenuation after 14 days treatment (Fig. 1). However, a higher dose (10 mg kg⁻¹) attenuates the hypothermic response after a single dose and this diminution of the response becomes marked after 14 days administration, decreasing somewhat 48 h after withdrawal of the drug (Fig. 1).

Tranlycypromine produces no alteration of the 8-OH-DPAT-induced temperature change after 14, 21 and 26 days administration but a major attenuation occurs by 35 days, which is still present after drug withdrawal (Fig. 1). The time required to alter this 5-HT₁ receptor-mediated response is the same as previously found necessary to decrease 5-HT₂ receptor number and the serotonin agonist-induced head-twitch behaviour¹¹.

Our next experiments tested the effect of repeated ECS on the response of mice to 8-OH-DPAT. After a course of one ECS daily for 10 days, ECS-treated mice show an attenuation of the 8-OH-DPAT-induced hypothermic response which reaches a nadir (15% of the control response) 6 days after the last treatment and returns to 60% of control values by 18 days after the last ECS (data not shown). When the ECS was given in a manner more similar to the clinical administration of electroconvulsive treatment (ECT), that is, 5 ECS spread out over 10 days, the attenuation of the 8-OH-DPAT-induced hypothermic response occurs more slowly but is still present 24 days after the last ECS (Fig. 2). Anaesthetic administration alone has no effect on the response (Fig. 2). Finally, a similar attenuation is seen 3 days after the last of a course of ECS given twice weekly for 4 weeks (data not shown).

The 8-OH-DPAT dose-hypothermic response curve is a conventional sigmoid⁷. Repeated ECS and amitriptyline shift this curve to the right in a similar fashion, suggesting that both the physical and pharmacological treatments produce a similar change in the response.

To date there has been little biochemical characterization of 5-HT₁ receptors in the mouse brain. However, their pharmacology clearly indicates that they do not have the same characteristics as those in rat brain^{6,7}. For example, (–) propranolol is a poor antagonist of the 8-OH-DPAT-induced hypothermic response in the mouse⁷, but a good antagonist in the rat⁶. The drug RU 24969 (5-methoxy-3-(1,2,3,6-tetrahydropyridin-4-yl)1H indole) which in the rat has been suggested to act at the 5-HT_{1B} receptor¹², does not induce hypothermia in rats¹³ or mice⁷, although it produces a striking motility syndrome in both species⁶. This drug is antagonized by (–)propranolol in the mouse but not in the rat⁶. Furthermore, preliminary results have indicated that the locomotor response of mice to RU 24969 is unchanged by the antidepressant regime reported here (G.M.G., R.J.D.S. and A.R.G., unpublished). Therefore, comparison of ligand-binding data in rats with these functional responses in mice may be inappropriate. Certainly, previous studies on 5-HT₁ receptors in rat brain tissue have indicated that in general, few changes occur¹⁴ except after administration of monoamine oxidase inhibitors which cause a decrease in the number of 5-HT₁ binding sites¹⁴. However, zimeldine, desipramine and imipramine have been reported to induce a two-component binding site consistent with 'down-regulation'¹⁵, and zimeldine (but not the tricyclic drugs) has been shown initially to depress the neuronal firing rate in the dorsal raphe nucleus, followed by a return to control values over 2 weeks suggesting serotonin autoreceptor desensitization¹⁶.

As it has been suggested that the 8-OH-DPAT-induced hypothermia results from an inhibition of serotonin release⁷, we asked whether the changes observed in our study could have resulted from an altered sensitivity of the postsynaptic serotonin receptor involved in the response, rather than an alteration in the presynaptic 5-HT₁ receptor. Our evidence, presented here, seems to exclude a 5-HT₂ postsynaptic change being responsible, as mianserin (2 mg per kg) produces a marked decrease in postsynaptic 5-HT₂ receptor number and function¹¹ but has no effect on the hypothermic response. Furthermore, repeated ECS treatment has been shown to produce the opposite effect on 5-HT₂ receptor number and function to other antidepressant treatments¹¹, but has the same effect on the 8-OH-DPAT-induced hypothermia. Nevertheless, the serotonin postsynaptic receptor involved in the temperature response has not yet been well characterized^{6,7}.

Repeated administration of various antidepressant drugs (tricyclics, mianserin, iprindole, zimeldine and monoamine oxi-

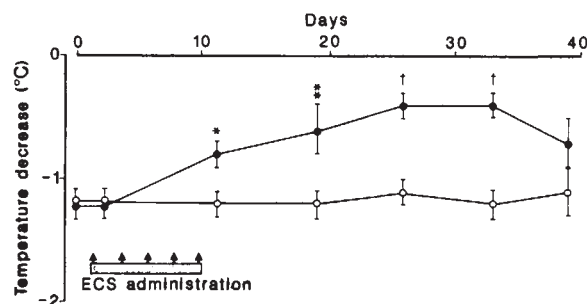


Fig. 2 Effect of repeated ECS on the hypothermic response of mice to 8-OH-DPAT. The graph shows mean $\pm$ s.e.m. (bars) temperature decrease in control (halothane anaesthesia, $\circ$) and ECS (90 V, 50 Hz, 1 s through earclip electrodes to halothane-anaesthetized mice, $\bullet$) 20 min after 8-OH-DPAT (0.5 mg kg⁻¹ s.c.). The ECS-treated group shows a significantly smaller temperature decrease than the control group as shown: * $P < 0.05$, ** $P < 0.005$, † $P < 0.001$. There is no difference in the basal temperature between the two groups at any time-point shown.

dase inhibitors) has been shown to decrease the density of the 5-HT₂ site in the frontal cortex of rats¹⁷ and mice¹¹ and also to decrease the 5-HT₂-mediated head-twitch behaviour¹¹. No simple hypothesis has been proposed to link this change in 5-HT₂ function to antidepressant activity because repeated ECS treatment, when given in a manner similar to the clinical administration of ECT, increases both 5-HT₂ receptor number and the 5-HT₂ receptor-mediated head-twitch behaviour¹¹. Our data, reported here, provide the first demonstration that these diverse treatments can produce the same functional change in a serotonin-mediated effect. Note that, first, the change is seen both during drug treatment and on withdrawal; second, the change develops fully only after long-term treatment; third, only the higher dose of mianserin produces the change and current recommendations¹⁸ favour higher doses than those originally used clinically; fourth, the highest doses of tricyclic antidepressants used here are unlikely to produce plasma levels in excess of those achieved clinically¹⁹; finally, repeated ECS produces a profound and long-lasting effect. Therefore, subsensitivity of the 5-HT₁ autoreceptor in specific brain regions may be a change induced by most antidepressant treatments and could serve as an important mechanism of their therapeutic action given the evidence for the involvement of serotonin in affective disorders²⁰.

G.M.G. holds a MRC Clinical Training Fellowship. We thank Astra, Ciba-Geigy, Organon and Smith Kline & French for the gifts of antidepressant drugs.

Received 18 July; accepted 19 August 1985.

- Peroutka, S. J. & Snyder, S. H. *Molec. Pharmacol.* **16**, 687-699 (1979).
- Pedigo, N. W., Yamamura, H. I. & Nelson, D. L. *J. Neurochem.* **36**, 220-226 (1981).
- Middlemiss, D. N. & Fozard, J. R. *Eur. J. Pharmacol.* **90**, 151-153 (1983).
- Gozlan, H., El Mestikawy, S., Pichat, L., Glowinski, J. & Hamon, M. *Nature* **305**, 140-142 (1983).
- Hamon, M. *et al. Eur. J. Pharmacol.* **100**, 263-276 (1984).
- Goodwin, G. M. & Green, A. R. *Br. J. Pharmacol.* **84**, 743-753 (1985).
- Goodwin, G. M., De Souza, R. J. & Green, A. R. *Neuropharmacology* (in the press).
- Ross, S. B. & Renyi, A. L. *Neuropharmacology* **16**, 57-63 (1977).
- Waldmeier, P. C., Greengrass, P. M., Baumann, P. & Maitre, L. *Postgrad. med. J.* **52**, Suppl. 3, 33-38 (1976).
- Baumann, P. A. & Maitre, L. *Naunyn-Schmiedeberg Archs Pharmacol.* **300**, 31-37 (1977).
- Goodwin, G. M., Green, A. R. & Johnson, P. *Br. J. Pharmacol.* **83**, 235-242 (1984).
- Cortes, R., Palacios, J. M. & Pazos, A. *Br. J. Pharmacol.* **82**, 202P (1984).
- Green, A. R., Guy, A. P. & Gardner, C. R. *Neuropharmacology* **23**, 655-661 (1984).
- Savage, D. D., Mendels, J. & Frazer, A. *Neuropharmacology* **19**, 1063-1071 (1980); *J. Pharmacol. exp. Ther.* **212**, 259-263 (1980).
- Fuxe, K., Ogren, S.-O., Agnati, L. F. & Calza, L. *Acta physiol. scand.* **114**, 477-480 (1982).
- Blier, P. & De Montigny, C. *J. Neurosci.* **3**, 1270-1278 (1983).
- Peroutka, S. J. & Snyder, S. H. *Science* **210**, 88-90 (1980).
- Grahame-Smith, D. G. & Aronson, J. K. *Oxford Textbook of Clinical Pharmacology and Drug Therapy* (Oxford University Press, 1984).
- Baumann, P. A., Gaillard, J.-M., Jonzier-Perey, M., Gerber, C. & Bouras, C. *Psychopharmacology* **84**, 489-495 (1984).
- Goodwin, F. K. & Post, R. M. *Br. J. clin. Pharmacol.* **15**, 393S-405S (1983).

Patterns of dye coupling in the imaginal wing disk of *Drosophila melanogaster*

Scott E. Fraser & Peter J. Bryant

Developmental Biology Center and Departments of Physiology and Biophysics and Developmental and Cell Biology, University of California at Irvine, Irvine, California 92717, USA

Responses of developing tissues to experimental disruption demonstrate that cell interaction is important both in generating positional information and in controlling growth¹. However, the mechanism by which cells interact and the range over which the interactions are effective are not known. In the imaginal disks of *Drosophila melanogaster*, experiments on pattern regulation following surgical ablation suggest that the cell interactions are very local in nature; in fact, most of the data can be explained by assuming that cells interact only with their immediate neighbours². In contrast, studies of cell division patterns in the same tissue indicate that the 'local' proliferative response to an ablation extends over a distance of up to about eight cell diameters³. Still longer-range interactions have been proposed on the basis of theoretical considerations⁴. It is possible that the interactions are mediated by the transfer of small molecules through gap junctions, as gap junctions are abundant in imaginal disks at the appropriate developmental stages⁵. We have explored the range, timing and directionality of dye coupling between the cells of the wing disk as a test of the possible role of gap junctions in imaginal disk patterning. Our results indicate that interactions over different ranges are possible depending on the nature of the molecule being transferred.

Dye coupling between the cells of the imaginal wing disk was monitored by the iontophoretic injection of either Lucifer yellow CH (ref. 6) or fluorescein complexon (Kodak) into one cell while observing the disk with an epifluorescence microscope. Both membrane potential measurements and direct observation of the injected dye were used to follow the progress of the dye injection and to ensure that the microelectrode tip remained within the impaled cell during the iontophoresis. The membrane potential typically reported by Lucifer yellow and fluorescein complexon-filled microelectrodes were -4 ± 2 mV and -7.5 ± 4 mV, compared with -13 ± 4 mV recorded by KCl-filled microelectrodes (all values: mean $\pm$ s.d.; $n = 25$). These differences are presumably caused by tip potential effects.

Injection of Lucifer yellow resulted in a characteristic pattern of dye passage, with dye spreading within a few seconds to fill a distinct small 'cluster' of cells with a sharp boundary (Fig. 1a). As the dye showed an affinity for nuclei, it was possible to count or estimate the number of cells per cluster. Clusters consisted of up to ~ 20 cells in disks of all ages investigated (2, 3, 4 and 5 days after oviposition at 25 °C). The cluster was often not symmetrical about the microelectrode tip, and further injection did not cause the dye to spread beyond the initial cluster. A second dye injection into another cell in a given cluster increased the fluorescence of the entire cluster, but did not lead to further dye spread. Thus, it is unlikely that the restriction observed was caused by either the dye being sequestered by the cells or the injected cells dying and thereby limiting the passage of the Lucifer yellow. Sequestration would not generate the same pattern after a second injection and cell death should uncouple the cells of the cluster from one another. An injection into a cell close to a dye-filled cluster resulted in a neighbouring cluster of dye-filled cells. Dye passage within a cluster does not seem to be caused by the presence of cytoplasmic bridges between cells as only ~ 5 -10% of imaginal disk cells are joined by such bridges (J. S. Ryerse, personal communication). Furthermore, when a large dye (fluorescein dextran) was injected, we found that only single cells or pairs of cells were filled (Fig.

3d), in agreement with previous work using horseradish peroxidase⁷. Although the boundaries to the Lucifer yellow-filled clusters were distinct and often rather straight, their positions do not seem to be reproducible, as demonstrated by comparing the cluster positions in either the left and right imaginal wing disks of the same animal or the disks of different animals (Fig. 1b).

Fluorescein complexon injection produced a smooth, graded pattern of dye passage, with no sharp boundaries to dye spread (Fig. 2). In young disks (2 days after oviposition), the dye spread to fill most or all of the epithelium. In 3-day disks, the dye filled a substantial fraction of the epithelium, with the edges of the filled region showing smooth gradients of fluorescence. In 4- and 5-day disks, the dye passage was also smooth and graded, but injection into the perimeter of the wing pouch resulted in predominantly circumferential spread of the dye about the wing pouch, with a much smaller radial spread (Fig. 2). The circumferential folding pattern of the disk epithelium could have contributed to this pattern by increasing the distance the dye must pass through the tissue for a given linear distance when viewed from above; however, the asymmetry of the dye passage seemed to exceed that expected from the distortion produced by folding. A second dye injection into a previously filled region of the disk epithelium resulted in additional spread of the complexon, with no noticeable boundaries to the smooth spread of the dye.

The reasons for the different patterns of passage of Lucifer yellow and fluorescein complexon are unclear, but probably involve either the different sizes or the different chemical natures of the two dyes. Although fluorescein complexon is heavier than Lucifer yellow (relative molecular mass (M_r) for ions 618 compared with 443), measurements made from CPK molecular models indicate that Lucifer yellow is at least 1 Å larger along its semi-major axis. Both dyes are negatively charged and seem to exist as divalent ions in physiological solutions according to vapour pressure osmometry (D. Woodbury, personal communication); however, their chemical properties are not entirely the same. Lucifer yellow has amino and sulphate groups exposed on its surface, whereas complexon has carboxyl groups exposed. Perhaps reflecting these chemical differences, Lucifer yellow demonstrated an affinity for cell nuclei in *Drosophila*, in contrast to fluorescein complexon which showed no such affinity. The difference in passage is unlikely to result from fluorescein complexon moving between cells through non-junctional membranes as the dye did not demonstrate noticeable membrane permeability when injected into the lumen or onto the surface of the disk. The lumen of the disk was injected with ~100 times the dye normally injected in our experiments and the margin of the disk was periodically examined for up to 30 min. The injected dye filled the entire lumen but left the epithelial cells, visible at the margin, unfilled (Fig. 3c). In contrast, intracellular dye fills showed fluorescence extending to the margin of the disk (Fig. 3a). When injected into different regions of larval salivary glands, fluorescein complexon revealed cell coupling only in those regions known to be coupled (based on Lucifer yellow injections and electrical coupling measurements). Fluorescein complexon is therefore a suitable membrane-impermeant dye for studying dye passage between insect cells.

Our analysis of the patterns of Lucifer yellow (248 successful injections) and fluorescein complexon (137 successful injections) passage demonstrate that there are no reproducibly positioned boundaries to dye passage, and that there is no correlation between the patterns of dye passage and the lineage restriction boundaries (compartment boundaries) described earlier^{8,9}. This conclusion is valid for both the dorso-ventral and antero-posterior lineage restriction boundaries, even though recent work indicates that the two restrictions are maintained by different mechanisms¹⁰. Our interpretation of the patterns of dye passage is different from that proposed by Weir and Lo^{7,11}, who concluded that there were defined boundaries to the dye passage, some of which coincided with compartment boundaries. The reasons for this difference are unclear, but are likely to be the product of the less stringent superposition criteria used

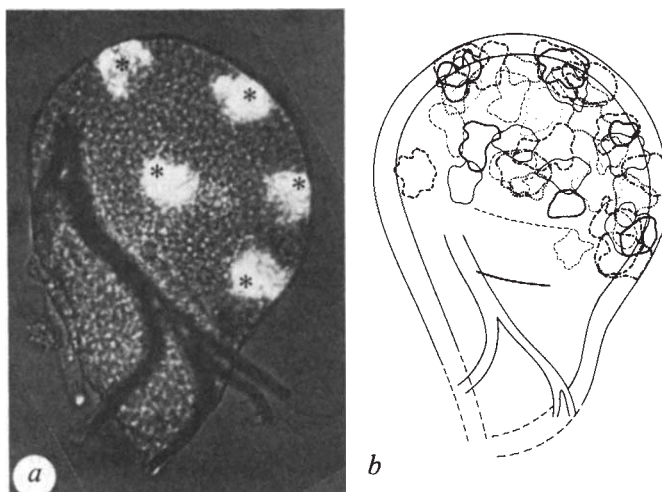
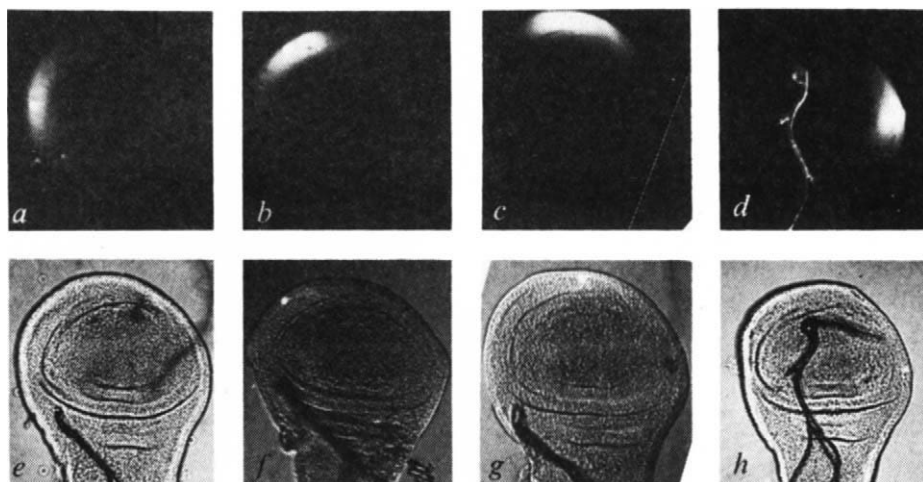


Fig. 1 Passage of Lucifer yellow in the imaginal wing disk. *a*, Pattern of dye passage in a 3-day-old imaginal wing disk following five sequential injections at different locations in one imaginal disk. The dye spread to a discrete group of cells from each injection, filling a cluster of about 20 cells (injected cell marked with an asterisk). This photograph was taken with a combination of phase and epifluorescent illumination so that the dye-filled clusters and the boundaries of the disk could be observed and compared. *b*, Superposition of the data from 39 dye injections in seven 3-day-old disks, demonstrating no consistent boundaries of the dye-filled clusters. Similarly, no consistent boundaries could be found in 2- or 4-day-old disks. In making comparisons of dye spread patterns, data were accepted only if the disk was undamaged and undistorted.

Methods. *a*, Imaginal disks were dissected from larvae of the OreRC strain in *Drosophila* Ringer's solution¹⁵ and allowed to settle onto a cleaned microscope slide, with the peripodial membrane touching the slide. The slide was mounted on the stage of a modified Zeiss UEM compound epifluorescence microscope and observed with a Zeiss filter set for Lucifer yellow fluorescence. The microscope was equipped with electronic shutters on phase and fluorescence light paths to minimize the exposure of the cells to wavelengths of light that could cause phototoxicity of dye-filled cells. The cells were impaled with thin-walled aluminosilicate microelectrodes (resistance 35–100 MΩ) using negative capacitance 'ringing' of the microelectrode or very brief depolarizing pulses of current. The tip shape and glass composition were both important for reliably recording a membrane potential with the dye-filled electrodes. The membrane potentials recorded by the dye-filled electrodes were used to confirm that the cell had been successfully impaled. The Lucifer yellow in the microelectrode (3% solution of the dilithium salt; Aldrich or Molecular Probes) was iontophoresed into the cells for 2–4 min, with 1–4 nA delivered as 250-ms pulses at 2 Hz, supplied by the current-passing microelectrode amplifier (NeuroData Instruments). The time constant of the microelectrode was fast enough to determine the membrane potential of the injected cells in the inter-pulse interval without the use of a bridge circuit. The membrane potential was assayed continuously with a digital storage oscilloscope (Gould), and the passage of the dye was monitored briefly at intervals with the epifluorescence microscope. *b*, The dye passage patterns were superimposed on a standard diagram of the disk, with alignment based on the anterior, posterior and distal margins as well as the no. 2 fold¹⁴. Photographs of individual disks were projected onto the diagram and slight size differences caused by developmental asynchrony were equalized by changing the magnification, using the antero-posterior dimension as the size criterion. As disk shape is highly reproducible, this method allowed accurate superposition of cluster outlines.

by Weir and Lo. Two different types of dye fill were reported by Weir and Lo^{7,11}; long-term (28–61 min) fills in which Lucifer yellow spread to fill a large fraction of the disk, and short-term (up to 5 min) fills in which either Lucifer yellow or carboxy-fluorescein spread to a relatively small group of cells. Although some of their short-term Lucifer yellow fills were shown histologically to be intracellular, this was not the case for many of the long-term Lucifer yellow fills used to support the claim of gap-junctional communication compartments. Membrane

Fig. 2 Patterns of dye passage after injection of fluorescein complexon into 4-day-old disks. *a-d*, Fluorescence; *e-h*, bright-field microscopy. The dye passed smoothly from the impaled cells, without the sharp boundaries observed with Lucifer yellow. Experimental conditions were as described in Fig. 1, except that the microelectrodes were filled with 100 or 200 mM fluorescein complexon (Kodak) and filter sets appropriate for fluorescein fluorescence were used. $\times 200$.



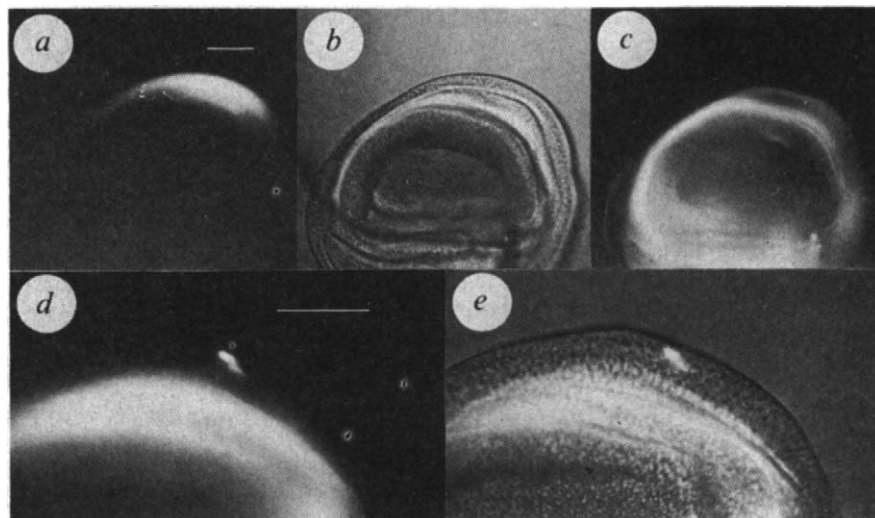
potentials, which can be used to verify the intracellular location of the electrode tip, were not recorded by Weir and Lo. Several of the long-term fills (Figs 1, 2 in ref. 11; Fig. 1 in ref. 7) do not reach the edge of the disk, but instead resemble the fluorescence patterns seen in our lumen fills (Fig. 3*c, e*). We have not found it possible to maintain stable impalements of the small, highly columnar imaginal disk cells for periods comparable with those claimed for their long-term fills^{7,11}. We therefore suggest that in the long-term fills much of the fluorescence is from dye that leaked into the disk lumen or into the space between the disk and the substrate, and that the apparent barriers to dye passage are caused by either natural or accidental folds in the disk epithelium. Weir and Lo also use their short-term fills to support their claim of communication compartments, but their data do not support the claimed consistency in positions of dye-fill boundaries.

The lack of correspondence between the compartment boundaries in the wing disk and the dye-passages contrasts with the finding that the insect body-segment boundary corresponds with a boundary to dye passage^{12,13}. It has been proposed that the insect segment represents a unit of repeated positional information; thus this physiological isolation of the segments from one another may have functional connotations. In contrast, the wing disk does not seem to possess repeated units of positional information, but instead behaves as a single morphogenetic field¹⁴.

Our results demonstrate the cell-to-cell transfer of two tracer dyes, and indicate that small molecules may travel over distances corresponding to several (Lucifer yellow) or many (fluorescein complexon) cell diameters in the imaginal disk. The discrete nature of the passage of Lucifer yellow to a cluster of up to 20 cells was not reflected in the longer-range passage of the slightly smaller dye, fluorescein complexon. The restriction of the Lucifer yellow spread could result from selectivity differences as suggested to explain the dye coupling restriction at the insect segment border^{12,13}; however, it remains possible that the restriction of the Lucifer yellow dye-passages is a function of its chemical nature.

Although much of the data on pattern formation and growth control in imaginal disks can be adequately explained by nearest-neighbour cell interactions¹, our results demonstrate that longer-range signalling is also possible. The two distinct distances over which dye passage was observed show some interesting similarities to those predicted by experimental and theoretical work. For example, the range of the proliferative response to a partial ablation of the disk³ and the size of the zone of non-proliferating cells at the dorso-ventral lineage restriction line¹⁰ are both very similar to the range of the Lucifer yellow dye passage. The differences in behaviour of the two dye molecules suggest that biologically important signalling molecules can be transferred through gap junctions over different ranges depending not only on the selective properties of the gap junction, but

Fig. 3 Fluorescent dye patterns following injection of fluorescein complexon or fluorescein dextran intracellularly or into the lumen of the disk. *a*, Intracellular injection of fluorescein complexon leads to a dye passage pattern that reaches the very edge of the disk epithelium as seen in bright-field microscopy (*b*). In contrast, injection of the dye into the lumen of the same disk (*c*) generates a smooth pattern that never reaches the margin of the disk, but instead remains one cell height from the margin. The cells visible at the margin of the disk remained unlabelled with the fluorescein complexon, indicating that the dye is a suitable membrane-impermeant tracer for cell-coupling determinations. The photograph in *c* was taken after the dye shown in *a* was intentionally bleached by exposure to the intense blue light of the fluorescence illuminator. *d, e*, Intracellular injection of fluorescein dextran (M_r 10,000; Molecular Probes), which is too large to pass through the gap junction channel, labels one or two cells (*d*, fluorescence; *e*, combined fluorescence and bright-field). The lumen of the disk was injected with fluorescein complexon for comparison. The fluorescein dextran fill demonstrates that dye passage through cytoplasmic bridges does not contribute to the patterns of dye spread observed following intracellular injection of either Lucifer yellow or fluorescein complexon. Comparison of the fluorescein dextran fill with the pattern of spread following injection of dye into the lumen of the disk further highlights the difference between the patterns observed from dye within the cells and in the extracellular space. Scale bar, 25 μm .



also on differences in affinity of the signalling molecules with their receptors and differences in the rates of synthesis and degradation of these molecules.

This investigation was supported by NIH grant HD06082 and NSF grants BNS 80-23638 and BNS 84-06307.

Received 5 July; accepted 6 August 1985.

- French, V., Bryant, P. J. & Bryant, S. V. *Science* **193**, 969-981 (1976).
- Held, L. L., Jr & Bryant, P. J. in *Pattern Formation: A Primer in Developmental Biology* (eds Malacinski, G. M. & Bryant, S. V.) 291-322 (Macmillan, New York, 1984).
- O Brochta, D. A. thesis, Univ. California, Irvine (1984).
- Meinhardt, H. in *Pattern Formation: A Primer in Developmental Biology* (eds Malacinski, G. M. & Bryant, S. V.) 47-72 (Macmillan, New York, 1984).
- Ryser, J. S. *Wilhelm Roux Arch. dev. Biol.* **191**, 335-339 (1982).
- Stewart, W. W. *Nature* **292**, 17-21 (1981).
- Weir, M. P. & Lo, C. W. *Dev. Biol.* **102**, 130-146 (1984).
- Garcia-Bellido, A., Ripoll, P. & Morata, G. *Nature new Biol.* **245**, 251-253 (1973).
- Brower, D. L., Lawrence, P. A. & Wilcox, M. *Dev. Biol.* **86**, 448-455 (1981).
- O Brochta, D. A. & Bryant, P. J. *Nature* **313**, 138-141 (1985).
- Weir, M. P. & Lo, C. W. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3232-3235 (1982).
- Warner, A. E. & Lawrence, P. A. *Cell* **28**, 243-252 (1982).
- Blennerhassett, M. G. & Caveney, S. *Nature* **309**, 361-364 (1984).
- Bryant, P. J. *J. exp. Zool.* **193**, 49-78 (1975).
- Schubiger, G. *Dev. Biol.* **26**, 277-295 (1971).

GTP-binding proteins couple cardiac muscarinic receptors to a K channel

Paul J. Pfaffinger*, Jennifer M. Martin†, Dale D. Hunter†, Neil M. Nathanson† & Bertil Hille*

* Department of Physiology and Biophysics and † Department of Pharmacology, University of Washington School of Medicine, Seattle, Washington 98195, USA

Binding of acetylcholine (ACh) to cardiac muscarinic ACh receptors (mAChR) activates a potassium channel that slows pacemaker activity¹⁻³. Although the time course of this activation suggests a multi-step process with intrinsic delays of 30-100 ms⁴⁻⁶, no second-messenger system has been demonstrated to link the mAChR to the channel. Changes in cyclic nucleotide levels (cyclic AMP and cyclic GMP) do not affect this K channel or its response to muscarinic agonists^{7,8}. Indeed, electrophysiological experiments argue against the involvement of any second messenger that diffuses through the cytoplasm⁹. We report here that coupling of the mAChR in embryonic chick atrial cells to this inward rectifying K channel requires intracellular GTP. Furthermore, pretreatment of cells with IAP (islet-activating protein from the bacterium *Bordetella pertussis*) eliminates the ACh-induced inward rectification. As IAP specifically ADP-ribosylates two GTP-binding proteins, N_i and N_o , that can interact with mAChRs¹⁰, we conclude that a guanyl nucleotide-binding protein couples ACh binding to channel activation. This represents the first demonstration that a GTP-binding protein can regulate the function of an ionic channel without acting through cyclic nucleotide second messengers.

Ionic currents of single atrial cells were studied using the whole-cell voltage-clamp technique^{11,12}, in which the recording pipette is sealed against the cell and the underlying membrane is broken by a pulse of suction, allowing the solution inside the pipette to equilibrate with the cytoplasmic contents. By using small embryonic cells and relatively large pipettes, we sought to achieve a more rapid equilibration than in previous work with adult atrial cells. Figure 1A shows typical ionic currents in a bathing solution containing 25 mM potassium recorded 5 min after the whole-cell voltage clamp was established. As in other cardiac cells¹³, hyperpolarizing voltage steps evoke large inward currents carried by K^+ ions, and depolarizing steps evoke only small outward K currents. The high conductance for inward K currents and low conductance for outward K currents is a characteristic sign of inward rectifying K channels. In these cells the K current is totally activated within a few milliseconds after the voltage step. In ventricular cells this background current is often termed I_{K1} ; however, in our atrial cells we will call it

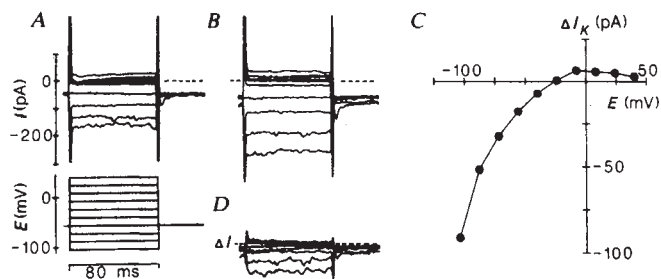


Fig. 1 Acetylcholine-induced currents in a single embryonic chick atrial cell. **A**, Whole-cell currents in 25 mM external K recorded under voltage-clamp conditions after equilibration with the pipette/intracellular solution for 5 min. The stable current level recorded from other cells varied considerably; however, the form of currents seen here is typical. The cell was held at -56 mV and currents were recorded during 80-ms voltage steps from -104 to +40 mV in 16-mV increments. Current records were recorded online onto a RACAL FM tape recorder, and were later filtered at 500 Hz and digitized into a LM² minicomputer. **B**, Whole-cell currents in the same cell recorded after the addition of 10 μM ACh. ACh was added immediately after the currents in **A** were recorded. Currents were recorded again after 1-2 min to determine the effect of ACh. Voltage steps and recording procedures were identical to those in **A**. **C**, ACh-activated current obtained by computer subtraction of the records in **A** from those in **B**. **D**, Current-voltage relation for the ACh-activated K current in the same cell. All records were analysed by computer for the steady-state current level at the end of the 80-ms voltage step.

Methods. Cells from 9-day-old embryonic chick atria were enzymatically dispersed²² and plated at a low density (5×10^5 cells per dish) in 35-mm tissue culture dishes to grow as single cells. After growth for 2-4 days at 37 °C in culture medium²² supplemented with 5% fetal calf serum, the single cells were spontaneously active. Beating could be arrested by addition of ACh. For most experiments, the external medium was changed to 25 mM K/avian Ringer's (100 mM NaCl, 25 mM KCl, 2.5 mM CaCl₂, 4 mM MOPS/NaOH, pH 7.3) at 21-25 °C. Single atrial cells with dimensions less than 15×40 μm were selected for electrical recordings via the whole-cell voltage-clamp technique^{11,12}. The use of small cells facilitated intracellular perfusion with pipettes of 2-4 MΩ resistance. The cells were allowed to equilibrate with pipette/intracellular solution (100 mM K aspartate, 20 mM KCl, 5 mM NaCl, 1 mM MgCl₂, 1 mM EGTA, 1.5 mM ATP, 100 μM GTP, 5 mM MOPS/KOH, pH 7.4) for 5 min before the experiment with ACh began. ACh (8 μl of 2 mM ACh dissolved in normal avian Ringer's) was added to the external solution to obtain a final concentration of 5-10 μM.

simply 'background K current' as two types of K channels are believed to contribute to it¹⁴. Addition of 10 μM ACh to the bath (Fig. 1B) increased currents by 50% in response to hyperpolarizing and depolarizing voltage steps; subtracting the records without ACh from those with ACh gives the ACh-induced component (Fig. 1C). Its steady-state current-voltage relation is steep for inward current and shallow for outward current (Fig. 1D). Like the background current in untreated cells, the ACh-induced current shows marked inward rectification.

We first show that the receptor involved is a mAChR and that the channel it activates is an inward rectifying K channel, as in cardiac cells from other species^{2,9,15-17}. Atropine is a selective antagonist of the muscarinic action of ACh. In all of five experiments atropine (10 μM) prevented the induction by ACh of additional K conductance. Several criteria were used to identify the type of channel involved. Figure 2 shows ACh-induced currents for different cells bathed in external solutions containing 2.5, 25 and 100 mM K. The zero-current potential shifts with external K concentration, as required for a K-selective channel, and the range of potential in which rectification occurs shifts by the same amount, a property unique to inward rectifying K channels. The ACh-induced current and the background K current have similar current-voltage relations and both are fully activated within a few milliseconds; both could be blocked by 1 mM Ba²⁺ or by 1 mM Cs⁺, as expected for inward rectifying

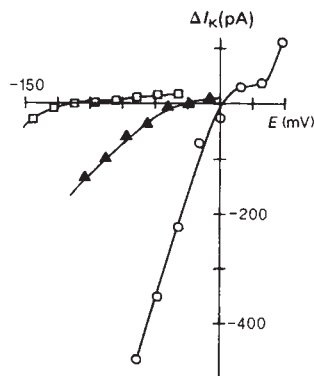


Fig. 2 Current-voltage relations for the ACh-activated K current recorded in three different cells bathed in external solution containing 2.5 mM ($\square$), 25 mM ($\blacktriangle$) or 100 mM ($\circ$) K. Measurements as in Fig. 1 except that the holding potential was adjusted to be near the expected reversal potential. All solutions were identical to those in Fig. 1 except for an equimolar substitution of KCl for NaCl in the external solutions to create the different K concentrations.

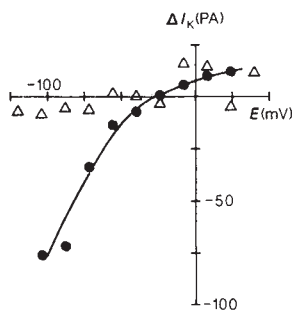


Fig. 3 The averaged current-voltage relations in 25 mM external K for the ACh-activated K currents in cells recorded either with 100 μ M GTP ($\bullet$, $n=6$) or without GTP (Δ , $n=6$) in the pipette/intracellular solutions. The experimental procedure was as in Fig. 1, except that we alternated between the normal pipette intracellular solution containing 100 μ M GTP and an identical one lacking GTP. Recordings were evaluated both for minimal response, as outlined in the text, and for the averaged response to 10 μ M ACh. Cells recorded without GTP were found to be significantly less responsive to ACh.

K channels¹³. Nevertheless, the ACh-induced and background K currents may involve different channels^{9,14}.

After finding a temporal correlation between the appearance in the embryo of muscarinic slowing of heart rate and changes in the isoelectric points of N-proteins, Halvorsen and Nathanson¹⁸ suggested that the GTP-binding protein N_i has a role in mAChR-mediated negative chronotropic response. To determine whether intracellular GTP is required for ACh-induced K currents, we made recordings from cells using pipettes containing either 100 μ M or no GTP. Of six cells recorded without GTP in the pipette, none showed any response to ACh addition, where our criterion for a minimal acceptable response was an increase in inward rectification with an inward slope conductance greater than 300 pS. All of six cells recorded with GTP in the pipette showed at least the minimal response. The average of the induced currents for these two groups of cells (Fig. 3) shows that in the absence of intracellular GTP, ACh fails to induce significant K current. The background K current (not shown) was present even without GTP in the pipette. In 15 cells recorded with GTP in the pipette, we monitored the background K current during the first 5 min after establishing the whole-cell voltage clamp; in four cells the K currents increased spontaneously during this period. In 10 of 11 cells with steady background K currents, K currents increased after application of ACh.

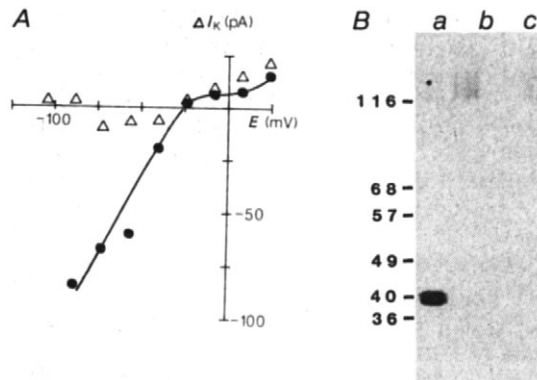


Fig. 4 ADP-ribosylation of GTP-binding proteins by IAP prevents the activation of inward rectifier K channels by ACh. **A**, Averaged current-voltage relation for the ACh-induced currents in 25 mM external K, recorded with 100 μ M GTP inside the pipette. Cells exposed for 24 h to IAP (Δ , $n=8$) were significantly less responsive to 10 μ M ACh. The averaged current for all untreated cells ($\bullet$, $n=9$) shows a clear inward rectification similar to that in Fig. 3. All experiments were as described in Fig. 1 except that in the case of the IAP-treated cells, the culture medium was exchanged for one containing 5–20 ng ml⁻¹ IAP, 24 h before the experiments began. No obvious damage to the cells or changes in the resting K currents could be seen in cells exposed overnight to IAP. **B**, Autoradiograph of a one-dimensional SDS-polyacrylamide gel of IAP-labelled membranes. Lane *a*, membranes from control cells, labelled with ³²P-NAD in the presence of IAP; *b*, membranes from control cells, labelled with ³²P-NAD in the absence of IAP; *c*, membranes from cells treated in culture with IAP (5 ng ml⁻¹) for 24 h, then labelled with ³²P-NAD in the presence of IAP. Cultured embryonic chick heart cells (50 μ g membrane protein per 35-mm dish) were incubated in the presence or absence of IAP (5 ng ml⁻¹) purified by the method of Sekura²³. After 24 h, cells were washed in phosphate-buffered saline (10 mM phosphate, pH 7.5), homogenized in 50 mM sodium phosphate buffer, pH 7.5, and resuspended in 100 mM Tris-HCl, pH 8.0. Membranes were labelled with IAP plus ³²P-NAD and analysed by SDS-gel electrophoresis and autoradiography¹⁸.

The dependence of K-channel induction on GTP could imply a novel requirement for GTP to be bound to the K channel, or it could mean that mAChR activation is communicated to the channel via a GTP-binding protein. For example, the muscarinic inhibition of adenylate cyclase is known to act through the GTP-binding protein N_i , and the toxin IAP, which ADP-ribosylates N_i , blocks this inhibition^{19,20}. We therefore set up an experiment to test for the involvement of an IAP-sensitive coupling protein. *In vitro* incubation of atrial cell membrane fractions with IAP and ³²P-NAD resulted in ADP-ribosylation of two polypeptides with relative molecular masses of 41,000 (41K) and 39K (Fig. 4B, lane *a*), corresponding to the α -subunits of the guanine nucleotide-binding proteins N_i and N_o , respectively^{18,21}. When cultures were pretreated with IAP, no polypeptides were labelled *in vitro* (Fig. 4B, lane *c*), indicating that this pretreatment of intact cells suffices to ADP-ribosylate all available N_i and N_o . We looked for ACh-sensitive K currents using pipettes containing 100 μ M GTP in cells pretreated with IAP; only one of eight cells showed a significant muscarinic response to ACh addition. The control cells showed responses above the minimal level in seven out of nine cells. The average induced currents for these two groups of cells (Fig. 4A) show that IAP treatment blocks the coupling of mAChR to inward rectifying K channels. The background K current (not shown) remained present in IAP-treated cells, and it did not increase during the first 5 min after GTP-containing pipettes were sealed onto the cell.

Our results demonstrate that the atrial muscarinic receptor interacts with an inward rectifying K channel through a GTP-binding protein such as N_i or N_o . Thus, in contrast to the nicotinic AChR in which a single protein molecule binds

neurotransmitter and also acts as an ion channel, the mAChR and the K channel it regulates appear to be distinct molecules. The involvement of GTP-binding proteins may be a common feature of all mAChR-mediated responses.

This work was supported by NIH grants NS08174, HL30639, GM07270, GM07750 and RR00374 and by a grant-in-aid from the American Heart Association, with funds contributed in part by the American Heart Association of Washington.

Received 6 June; accepted 12 August 1985.

- Loewi, O. & Navratil, E. *Pflügers Arch. ges. Physiol.* **214**, 678–688 (1926).
- Trautwein, W. & Dudel, J. *Pflügers Arch. ges. Physiol.* **266**, 324–334 (1958).
- Giles, W. R. & Noble, S. J. *J. Physiol., Lond.* **261**, 103–123 (1976).
- Hill-Smith, I. & Purves, R. D. *J. Physiol., Lond.* **279**, 31–54 (1978).
- Osterrieder, W., Yang, Q.-F. & Trautwein, W. *Pflügers Arch. ges. Physiol.* **389**, 283–291 (1981).
- Nargeot, J. *et al.* *J. gen. Physiol.* **79**, 657–678 (1982).
- Trautwein, W., Taneguchi, J. & Noma, A. *Pflügers Arch. ges. Physiol.* **392**, 307–314 (1982).
- Nargeot, J., Nerbonne, J. M., Engels, J. & Lester, H. A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2395–2399 (1983).
- Soejima, M. & Noma, A. *Pflügers Arch. ges. Physiol.* **400**, 424–431 (1984).
- Florio, V. A. & Sternweis, P. C. *J. biol. Chem.* **260**, 3477–3483 (1985).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).
- Matteson, D. R. & Armstrong, C. M. *J. gen. Physiol.* **83**, 371–394 (1984).
- Noble, D. *J. Physiol., Lond.* **353**, 1–50 (1984).
- Sakmann, B., Noma, A. & Trautwein, W. *Nature* **303**, 250–253 (1983).
- Giles, W. R. & Noble, S. J. *J. Physiol., Lond.* **261**, 103–123 (1976).
- Noma, A. & Trautwein, W. *Pflügers Arch. ges. Physiol.* **377**, 193–200 (1978).
- Momose, Y., Giles, W. & Szabo, G. *Biophys. J.* **45**, 20–22 (1978).
- Halvorsen, S. W. & Nathanson, N. M. *Biochemistry* **23**, 5813–5821 (1984).
- Hazeki, O. & Ui, M. *J. biol. Chem.* **256**, 2856–2862 (1981).
- Kurose, H. & Ui, M. *J. Cyclic Nucleotide Protein Phosphoryl. Res.* **9**, 305–318 (1983).
- Sternweis, P. C. & Robishaw, J. D. *J. biol. Chem.* **259**, 13806–13813 (1984).
- Nathanson, N. M. *J. Neurochem.* **41**, 1545–1549 (1983).
- Sekura, R. D., Fish, F., Manclark, C. R., Meade, B. & Zhang, Y. *J. biol. Chem.* **258**, 14647–14651 (1983).

Uncoupling of cardiac muscarinic and β -adrenergic receptors from ion channels by a guanine nucleotide analogue

Gerda E. Breitwieser & Gabor Szabo

Department of Physiology and Biophysics, The University of Texas Medical Branch, Galveston, Texas 77550, USA

Guanine nucleotide binding proteins, interchangeably called N or G proteins, seem to be the primary signal-transducing components of various agonist-induced cell membrane functions^{1–3}. In the heart, G proteins have been implicated in β -adrenergic modulation of the slow inward Ca^{2+} current. We have investigated the role of G proteins in muscarinic activation of an inwardly rectifying, acetylcholine (ACh)-induced K^+ current (I_{ACh}), and β -adrenergic activation of an (isoprenaline)-induced Ca^{2+} current (I_{si}). Here we report that intracellular application of the non-hydrolysable GTP analogue 5'-guanylylimidodiphosphate (GppNHP) brought about an agonist-induced, antagonist-resistant, persistent activation of I_{ACh} and I_{si} . This functional uncoupling of channel from receptor suggests that the muscarinic receptor and the I_{ACh} channel are separate molecular structures. Membrane conductance responses to sequential activation of muscarinic and β -adrenergic receptors demonstrate that in contrast to the muscarinic inhibition of I_{si} , muscarinic stimulation of I_{ACh} is mediated by a G protein via a pathway that does not involve adenylate cyclase. Taken together, the results support the notion that agonist is required to induce GppNHP binding and/or activation of the G proteins. Once triggered by agonist, the control system remains maximally activated, thereby transforming the cell so that it no longer responds to subsequent homologous receptor-mediated signals.

The broken-patch technique⁴ was used to voltage-clamp single atrial cells obtained by enzymatic dissociation of bullfrog atria⁶. Figure 1a shows that, in agreement with previous observations⁷, bath application of 1 μM ACh evokes an inwardly rectifying

K^+ current (I_{ACh}) which is fully blocked by 0.5 μM atropine, a muscarinic antagonist. I_{ACh} dissipates on removal of ACh; the membrane current returns to control levels following a 4–6 min superfusion with ACh-free Ringer's solution, and the cell responds to repeated applications of ACh (data not shown).

As binding of agonists to muscarinic receptors is modulated by guanine nucleotides⁸, it is reasonable to suppose that muscarinic activation of I_{ACh} is mediated by a G protein. This hypothesis was tested by examining the effects of intracellular GppNHP on the development of I_{ACh} . Bath application of ACh to a cell loaded with GppNHP through the patch pipette also evokes an inwardly rectifying current (Fig. 1b). The current-voltage (I - V) relationship for this current is similar to that observed in the absence of GppNHP, and is therefore identified as I_{ACh} . However, in GppNHP-treated cells I_{ACh} is not blocked by application of atropine, nor does I_{ACh} dissipate on removal of ACh; superfusion of the cells for up to 15 min with ACh-free Ringer's solution had no effect on I_{ACh} (data not shown). These results demonstrate that guanine nucleotides are involved in the muscarinic receptor-induced activation of I_{ACh} , and suggest the involvement of a G protein. The results also demonstrate that GppNHP functionally uncouples the I_{ACh} channel from the muscarinic receptor, since atropine or removal of agonist does not decrease I_{ACh} in the presence of GppNHP. It is tempting to conclude that functional uncoupling also reflects a physical uncoupling, and that the muscarinic receptor and the inwardly rectifying K^+ channel responsible for I_{ACh} reside on distinct proteins. There is indirect physiological and biochemical support for this contention. For example, in embryonic chick heart, muscarinic receptors appear before guanine nucleotide-sensitive binding, and functional coupling to adenylate cyclase or I_{ACh} is absent until guanine nucleotide sensitivity is achieved⁹. Our results support the notion that a GTP binding protein, which we denote here as G_K , mediates the muscarinic activation of the I_{ACh} . GppNHP maintains the conductive channel state, implying that under normal physiological circumstances, hydrolysis of GTP is necessary to return the channel to an agonist-responsive state. Note that in frog heart, GppNHP neither alters binding capacity nor causes irreversible binding of ACh¹⁰.

β -Adrenergic activation of I_{si} is known to be mediated by a G protein, G_s , interacting with the adenylate cyclase⁴. For comparison with the results obtained for I_{ACh} , the effects of GppNHP on β -adrenergic receptor activation of I_{si} are shown in Fig. 1c, d. In the absence of GppNHP (Fig. 1c) bath application of 1 μM isoproterenol causes a large increase in I_{si} , which can be fully blocked by application of 1 μM propranolol, a β -adrenergic antagonist. While there is a slight 'run-down' of I_{si} (see Fig. 1 legend), superfusion of the cell with isoproterenol-free Ringer's rapidly (within 3–5 min) reduces I_{si} to control values (data not shown). With GppNHP in the pipette (Fig. 1d) isoproterenol also induces increases in I_{si} , which, in contrast, are not blocked by propranolol or reduced by removal of agonist. Biochemical studies of frog erythrocyte membranes have demonstrated an isoproterenol-induced increase in cyclic AMP production which, in the presence of GppNHP, was not inhibited by propranolol or reversed by dilution of the assay medium¹¹. Our observation of a persistent, propranolol-resistant activation of I_{si} in the presence of GppNHP reflects an irreversible activation of adenylate cyclase via G_s , leading to increased levels of calcium channel activation.

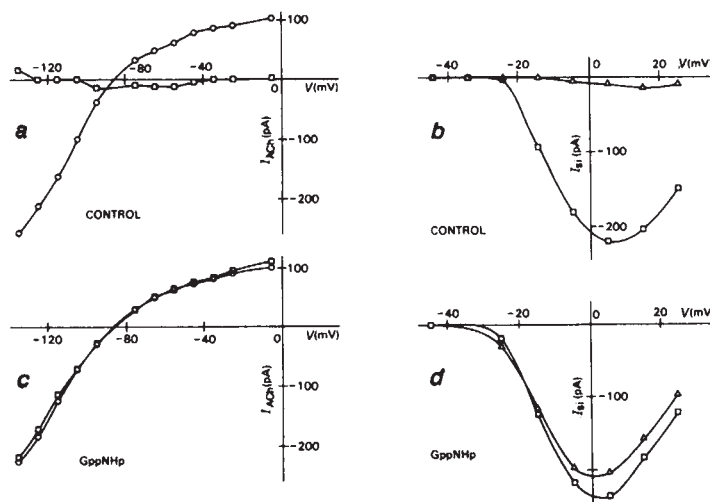
Intracellular GppNHP alone (for example, 0.1 mM GppNHP in the pipette without agonist in the bathing solution) had no effect on the membrane currents during the time course of our experiments, up to 30 min. Formation of the activated G protein must therefore be preceded by stimulation of the appropriate receptor. This observation is consistent with the behaviour of G proteins in cell-free preparations^{2,3}. Furthermore, if we assume that the activated form of the G protein is G-GppNHP, our results also imply that GDP-GppNHP exchange is negligibly slow.

Muscarinic receptors interact with a pertussis toxin-sensitive, inhibitory, guanine nucleotide binding regulatory protein, G_i ,

Fig. 1 Effect of intracellular GppNHP on ACh-induced inwardly rectifying K^+ current (I_{ACh}) and isoproterenol-induced slow inward current (I_{si}). I_{ACh} (a, b) was calculated by subtracting the control currents (no drug) from the current measured in the presence of $1 \mu M$ ACh ($\circ$) or of $1 \mu M$ ACh plus $0.5 \mu M$ atropine ($\square$). The currents were taken at 200 ms to minimize contamination from transient currents and the slow outward K^+ current¹². For these experiments, $0.5 mM$ $CdCl_2$ was added to the standard bathing solution to block I_{si} . a, A typical I_{ACh} -V pair (from 20 similar pairs) for a cell held with a control pipette containing no GppNHP. b, A typical I_{ACh} -V pair (from seven similar pairs) for a cell held with a GppNHP-containing pipette. I_{si} (c, d) was calculated by subtracting the control peak inward current, small in the absence of drugs, from the peak inward current measured in the presence of $1 \mu M$ isoproterenol ($\square$) or of $1 \mu M$ isoproterenol plus $1 \mu M$ propranolol (Δ). For these experiments, $0.2 mM$ $BaCl_2$ was added to the standard bathing solution to block K^+ currents. c, A typical I_{si} -V pair (from three similar pairs) for a cell held with a control pipette. d, A typical I_{si} -V pair (from three similar pairs) for a cell held with a GppNHP-containing pipette. There is an observable run-down of the isoproterenol-induced I_{si} in the absence or presence of GppNHP, $t_{1/2} \sim 12$ -15 min, which accounts for the slight decrease in I_{si} after addition of propranolol in the presence of GppNHP. This probably reflects a depletion of cellular ATP as a result of constant adenylate cyclase stimulation, as repeated brief pulses of isoproterenol followed by washout can elicit the same peak stimulation. Control experiments performed in the absence of Ba^{2+} or Cd^{2+} produced similar results, the analysis of which is rather complex. Therefore, we preferred to use the selective channel blockers.

Methods. Single myocytes were prepared from atria of the bullfrog *Rana catesbeiana* by enzymatic dissociation⁶. The whole-cell patch-clamp technique⁵ was used to establish electrical contact with cell interior. The patch pipettes were fabricated from square-bore borosilicate glass using a single pull, and were filled with $100 mM$ KCl (control pipettes), or $100 mM$ KCl, $0.1 mM$ GppNHP (Sigma), $1 mM$ HEPES pH 7.4 (GppNHP pipettes). For the range of pipette resistances selected (10 - $14 M\Omega$), the exchange of solute between pipette and cell interior was sufficient to produce a saturation of the nucleotide response, yet it was slow enough not to affect native Ca^{2+} channel activity or responsiveness to β -adrenergic agonists for at least 45 min. One drawback of high pipette resistance was poor frequency response (3 -dB cut-off near $300 Hz$), which was nevertheless sufficient to correctly estimate both I_{ACh} and I_{si} . Cell membrane currents flowing through the pipette were measured with a current-to-voltage converter connected to the pipette by way of Ag/AgCl electrodes. Membrane potential was controlled by a programmable waveform generator connected to the solution bathing the cell via an Ag/AgCl electrode. Cells were held near the resting potential, at $-85 mV$. I - V relationships were obtained from records of the membrane current in response to test pulses to various potentials ($500 ms$) followed by rest periods ($5 s$) at $-85 mV$. The standard HEPES Ringer's solution contained (in mM): $90 NaCl$, $5 MgCl_2$, $2.5 KCl$, $2.5 CaCl_2$, 10 glucose, 20 HEPES, pH adjusted to 7.4 with $NaOH$, and $5 \mu M$ tetrodotoxin to block Na^+ currents. All experiments were carried out at room temperature (20 - $22^\circ C$).

whose activation counteracts the β -adrenergic stimulation of the adenylate cyclase³. This mechanism explains the inhibitory effect of ACh on isoproterenol-induced increases of I_{si} . If, as we have observed, GppNHP permanently activates a specific G protein only after stimulation of a specific receptor by agonist, then the β -agonist-induced, apparently persistent I_{si} observed in the presence of GppNHP should be inhibited by ACh in a persistent, muscarinic antagonist-independent manner, creating a state in which ionic currents are no longer responsive to agonists or antagonists. A test of this prediction is illustrated in Fig. 2. The upper row of traces was obtained in the absence of GppNHP in the pipette. Addition of isoproterenol causes the expected increase in I_{si} which can be blocked by addition of ACh (in the continued presence of isoproterenol). Following the removal of both agonists, the cell responds to a second application of isoproterenol, which can again be inhibited by ACh. These effects are presumably mediated via interactions of G_s and G_i with adenylate cyclase. In the presence of GppNHP (lower row of traces in Fig. 2), isoproterenol also causes an increase in I_{si} , which can be blocked by addition of ACh (as in the control condition). In contrast to the control condition, however, the inhibitory effect of ACh is permanent; removal and re-application of isoproterenol or isoproterenol plus ACh causes no further changes in current. It seems that neither G_s nor G_i binds significant amounts of GppNHP until the appropriate receptor is stimulated by agonist. Thus, addition of isoproterenol results in a condition identical to that of Fig. 1d, in which an irreversibly activated G_s -GppNHP complex causes sustained activation of adenylate cyclase. G_i makes no contribution to the activity until the cell is exposed to muscarinic agonist. Addition of ACh then, and only then, allows irreversible activation of G_i -GppNHP, resulting in a decreased I_{si} through inhibition of the adenylate cyclase. In this case, however, removal of agonists has no effect on the current, as both G_i -GppNHP and



G_s -GppNHP are irreversibly uncoupled from their respective receptors. Addition of isoproterenol to a cell previously exposed to ACh in the presence of intracellular GppNHP causes no stimulation of I_{si} , suggesting that the presence of G_i -GppNHP precludes G_s stimulation of I_{si} .

The ability of β -adrenergic receptors to affect muscarinic receptor-mediated activation of I_{ACh} through the same system of interactions was tested using a protocol similar to that of Fig. 2. In this case, however, addition of $1 \mu M$ isoproterenol in the presence of $1 \mu M$ ACh had no effect on I_{ACh} in either the presence or absence of GppNHP (data not shown); this implies

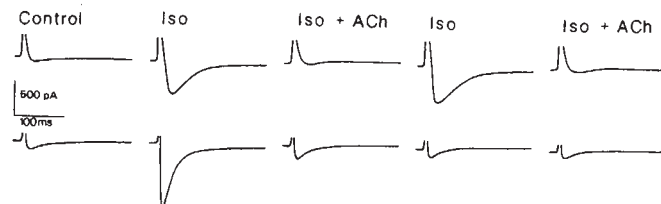


Fig. 2 Effect of intracellular GppNHP on the muscarinic modulation of I_{si} induced by β -adrenergic receptor stimulation. Upper trace, control pipette. Lower trace, GppNHP pipette. Voltage-clamp records were obtained using the methods described for Fig. 1. Cells were held at $-85 mV$ and depolarized to $-5 mV$. K^+ currents were blocked by $0.2 mM$ $BaCl_2$ in the standard Ringer's solution, which also contained: no drug (Control), $1 \mu M$ isoproterenol (Iso), $1 \mu M$ isoproterenol plus $1 \mu M$ ACh (Iso + ACh). The first application of isoproterenol + ACh was followed by a 10-min superfusion with control Ringer's solution. Each set of traces represents an experiment on an individual cell. Other experiments (4 for control pipettes, 10 for GppNHP pipettes) gave similar results. Note that the persistent block of I_{si} following the first application of ACh develops only in the presence of GppNHP in the cell.

that alterations of the adenylate cyclase activity by G_i and G_s do not affect I_{ACh} , and that I_{ACh} must therefore be regulated by a different mechanism that does not involve G_s or the adenylate cyclase. Whether G_K , which mediates the I_{ACh} response, is identical to G_i , remains to be established.

This work was supported by NIH grant HL24820 and by the American Heart Association, Texas Affiliate. We thank Drs A. Otero and A. M. Brown for helpful discussions.

Received 28 May; accepted 2 August 1985.

1. Rodbell, M. *Nature* **284**, 17-22 (1980).
2. Gilman, A. G. *Cell* **36**, 577-579 (1984).
3. Birnbaumer, L. *et al. Recent Prog. Horm. Res.* **41** (in the press).
4. Sperelakis, N. *Membrane Biochem.* **5**, 131-166 (1984).
5. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1981).
6. Hume, J. R. & Giles, W. J. *gen. Physiol.* **78**, 18-43 (1981).
7. Momose, Y., Giles, W. & Szabo, G. *Biophys. J.* **45**, 20-21 (1984).
8. Galper, J. B. & Ganz, P. in *Receptor Science in Cardiology* (eds Haft, J. I. & Karliner, J. S.) (Futura, New York, 1984).
9. Galper, J. B., Dziekan, L. C. & Smith, T. W. *J. biol. Chem.* **259**, 7382-7390 (1984).
10. Burgisser, E., De Lean, A. & Lefkowitz, R. J. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1732-1736 (1982).
11. Pappano, A. I. & Inoue, D. *Fedn Proc.* **43**, 2607-2612 (1984).
12. Hume, J. R. & Giles, W. J. *gen. Physiol.* **81**, 153-194 (1983).

Sodium-activated potassium current in cultured avian neurones

C. R. Bader, L. Bernheim & D. Bertrand

Department of Physiology, Centre Medical Universitaire, 1211 Geneva 4, Switzerland

An important characteristic of excitable cells is their ability to repolarize rapidly after a depolarizing event such as a sodium action potential. This ability is caused by several fast potassium currents such as I_A and I_K (ref. 1) and possibly also a fast calcium-dependent potassium current ($I_{K(CN)}$)². Suppression of only one of the potassium channels by mutation or drugs can lead to anomalous behaviour in the animal² or to epileptic discharges^{3,4} in *in vitro* preparations. Fast potassium currents vary in their inactivation properties¹, and this may be the reason why many types of potassium currents are required for the adequate functioning of neurones in various physiological conditions. Here we describe a neuronal potassium current that is activated by intracellular sodium. This current can probably be triggered by the sodium entering the neurone during a single action potential and shows little voltage inactivation. It could therefore participate in the fast repolarization process in situations where other potassium currents are partially inactivated. Because this current will be activated by any process leading to sufficient entry of sodium, it may also be triggered by sodium-mediated excitatory synaptic events. This could help to control the level of excitability in response to synaptic inputs. A sodium-dependent potassium current was recently described in heart cells⁵, but it was suggested that this may only become activated in pathological conditions.

Ciliary and trigeminal ganglia were dissected from 7-8-day-old chick or quail embryos. The dissociation and culture methods were similar to those described previously⁶. Dissociated neurones were studied with the whole-cell recording method⁷ using patch electrodes⁸. Figure 1 shows the effect of superfusing a neurone with 1 μ M tetrodotoxin (TTX); TTX suppressed an early inward current but, somewhat surprisingly, also blocked an outward current. Note that the peak of the outward current decreased in parallel with the decrease in inward current, suggesting a similar sensitivity to TTX. The reversal potential of the TTX-suppressible outward current was measured at the time indicated by the arrowhead in Fig. 1b. In four cells the reversal potential was -73 ± 5 mV (mean $\pm$ s.d.), a value close to the potassium equilibrium potential calculated from the known extra- and intracellular potassium concentrations (we assume

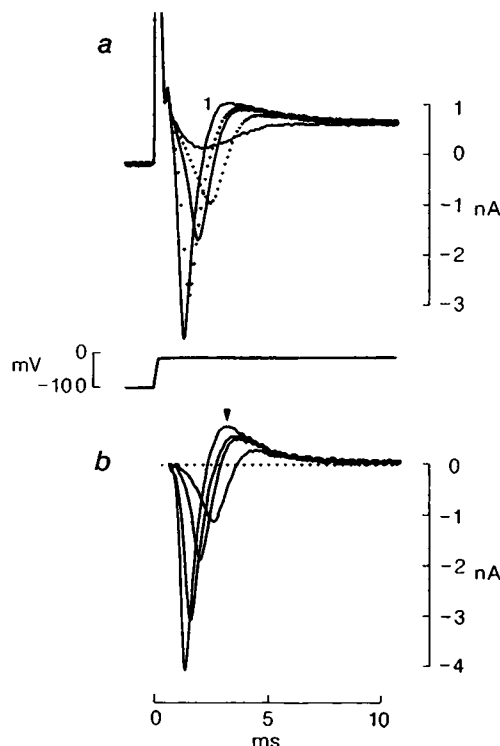


Fig. 1 Inward and outward currents suppressible by TTX. Quail trigeminal neurone cultured for 2 days. **a**, Progressive block by 1 μ M TTX and recovery in control solution. The voltage was stepped every 5 s from a holding level of -100 mV to -10 mV for 20 ms and the corresponding current was recorded. The continuous line labelled 1 is the current recorded before addition of TTX; the two other continuous lines represent a partial and total block by TTX (1 min after the superfusion with TTX). The dotted traces were recorded during recovery from the TTX-induced block. Inward current is plotted negatively. **b**, The current trace recorded in the full TTX-induced block was subtracted from all other current traces in **a**. The arrowhead indicates the time at which the voltage was stepped to other values when reversal potential measurements were performed (see text).

Methods. Patch electrodes were filled with (mM) K-acetate (130), KCl (10), NaCl (10), EGTA (5, but 20 mM in this figure), HEPES (5), glucose (5); the pH was adjusted to 7.4 with KOH. In several experiments (including that shown here), 5 mM Na_2ATP was added to the solution (NaCl was then omitted). The control superfusion solution was (mM) NaCl (150), KCl (5.4), CoCl_2 (3), HEPES (5), MgCl_2 (0.8), glutamine (2), glucose (16.7), bovine serum albumin (20 $\mu\text{g ml}^{-1}$), minimum essential medium vitamins, penicillin (100 U ml^{-1}), streptomycin (1 mg ml^{-1}); pH 7.3. The recording and storage procedures were as described previously⁸.

that the intracellular potassium concentration is in equilibrium with that of the solution in the patch pipette^{9,10}). Changing the potassium concentration displaced the reversal potential of the TTX-sensitive outward current, as expected for a current carried primarily by potassium ions.

Additional evidence that potassium is a major charge carrier for the TTX-suppressible outward current stemmed from the observation that it can be blocked by a combination of tetraethylammonium (TEA) and 4-aminopyridine (4-AP), two agents known to block potassium channels. This blocking effect of TEA and 4-AP allowed us to subdivide the total current suppressible by TTX into a sodium and a potassium component (Fig. 2).

In view of the high specificity of TTX for the sodium channel¹⁶ and the close parallelism in the reduction of potassium and sodium currents in the presence of TTX, we supposed that the potassium current was linked to the entry of sodium into the neurone. Several facts supported this idea. Extracellular sodium substitution with either choline or Tris suppressed both the sodium current and the potassium current within ~ 2 min. When

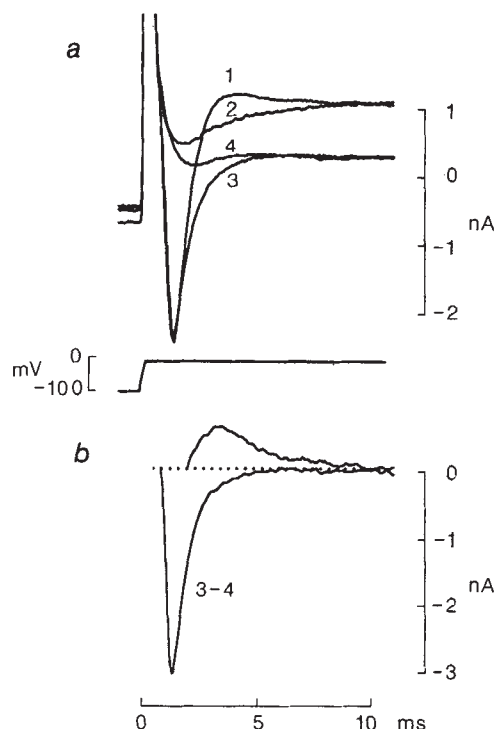


Fig. 2 Visualization of the potassium current suppressible by TTX. Quail trigeminal ganglion neurone cultured for 2 days. *a*, Currents were recorded during a step from -100 to -10 mV. Trace 1 is a recording in the presence of control medium; trace 2, in the presence of $1 \mu\text{M}$ TTX; trace 3, in the presence of 4-AP (4 mM) and TEA (20 mM); trace 4, in the presence of TTX, 4-AP and TEA. Note that the difference between traces 2 and 4 represents the voltage-dependent potassium currents, which are also present in these neurones (C.R.B. and D.B., unpublished data). *b*, The sodium current without interference by the potassium current was calculated by subtracting trace 4 from trace 3 (label 3-4). The potassium current suppressible by TTX is obtained by subtracting the sodium current (3-4) from the difference between traces 1 and 2. In this cell, the peak conductance of the potassium current was 10 nS . In 30 cells, the mean conductance was $8.7 \pm 3.5 \text{ nS}$.

partial inactivation of the sodium current was induced, a corresponding decrease of the potassium current was observed. When the voltage of a neurone was stepped to levels more depolarized than the sodium equilibrium potential (E_{Na}), no potassium current was seen (Fig. 3). We also observed that the duration of the potassium current was increased by a procedure that prolonged the sodium current. When *N*-bromoacetamide (NBA), which reduces the inactivation of the sodium current^{11,12}, was added to the solution in the pipette (1 mM), 7–10 min after breaking the patch, the sodium current was prolonged and the duration of the potassium current increased. Whereas, under normal conditions, the potassium current decayed to zero within $10 \pm 3 \text{ ms}$ ($n = 23$), in the presence of NBA the current was still activated to more than 60% of its maximal value after 20 ms.

Given the nature of the cells studied, that is, neurones with their processes, we wondered whether inadequate voltage control could account for the observation of the TTX-suppressible potassium current. Because it is almost impossible to prove that a neurone with its neurites is adequately voltage clamped, we examined neurones 3–5 h after dissociation. At that time, a neurone consists of a soma with no processes, but it already adheres firmly to the Petri dish and possesses voltage-dependent sodium and potassium currents. First, we ensured that our one-electrode voltage-clamp system was efficiently space-clamping these neurones by recording with a second-patch electrode located as far away as possible from the electrode connected to the voltage-clamp system. The voltages recorded by the two electrodes during steps in voltage clamp were identical.

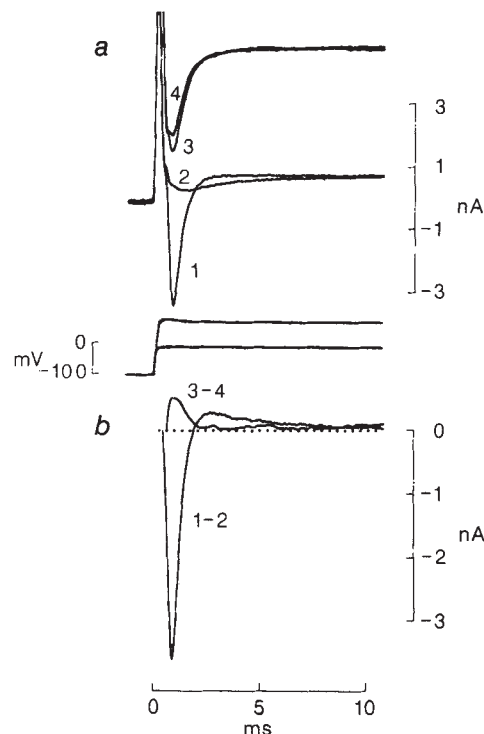


Fig. 3 The potassium current suppressible by TTX depends on sodium entry. Quail trigeminal ganglion neurone cultured for 3 days. *a*, The voltage of a neurone was stepped first from -100 mV to -10 mV (trace 1) and then to $+70$ mV (trace 3). TTX was then added to the medium and traces 2 and 4 were recorded at -10 and $+70$ mV, respectively. *b*, Differences between traces obtained at a given voltage in the presence or absence of TTX. When the sodium current reversed its polarity (3-4), the potassium current was absent. The reversal potential for the sodium current was at $+63 \text{ mV}$ in this cell; the calculated E_{Na} was $+68 \text{ mV}$.

We conclude that these cells without neurites are adequately space-clamped. Recordings similar to those illustrated in Fig. 1 were observed in these cells. Therefore, the TTX-suppressible potassium current discussed here cannot be explained by inadequate voltage control.

Finally, we performed an experiment in which the putative sodium-sensitive outward current was turned on without activating the voltage-dependent sodium current. Because we had observed that the potassium current suppressible by TTX was sensitive to 4-AP and TEA (see Fig. 2), we tested whether these drugs would also suppress a constant potassium current activated by a high intracellular sodium concentration. The voltage range chosen for this test was in a range where voltage-dependent potassium currents would not be activated (-130 to -60 mV). The presence of 50 mM sodium in the pipette caused persistent activation of a current that could be suppressed by 4-AP and TEA (Fig. 4). Because it has recently been shown that 4-AP and TEA can reduce a H^+ current¹³, it was necessary to demonstrate that the persistent current was carried by potassium ions. We found that the reversal potential for this current was near the theoretical E_{K} in two conditions, E_{K} at -24 mV (see Fig. 4) and E_{K} at -110 mV (data not shown). TEA and 4-AP had no effect when the sodium salts in the pipette were replaced with choline chloride and Tris-ATP, indicating that the potassium current suppressible by 4-AP and TEA in Fig. 4 was indeed activated by intracellular sodium.

We conclude that the potassium current described here is triggered by an increase in the intracellular sodium concentration ($[\text{Na}]_{\text{in}}$). Can it be called $I_{\text{K}(\text{Na})}$ or does sodium indirectly activate another neuronal potassium current such as $I_{\text{K}(\text{Ca})}$ (ref. 14)? Activation of $I_{\text{K}(\text{Ca})}$ through a sodium induced rise in $[\text{Ca}]_{\text{in}}$ seems unlikely. All experiments were performed in the absence of extracellular calcium, and the solution in the pipette contained the calcium chelator EGTA at a concentration of

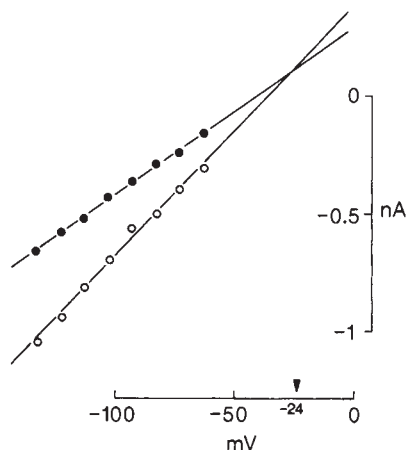


Fig. 4 Effect of a high intracellular sodium concentration. Quail trigeminal neurone cultured for 3 days. Superfusion with a solution containing 110 mM NaCl and 40 mM KCl. The solution in the pipette was (mM) NaCl (40), Na_2ATP (5), EGTA (5), potassium acetate (70), KCl (20), KOH (13), HEPES (5), glucose (5). The voltage of the neurone was held at -100 mV and stepped every 5 s for 30 ms to one of several voltages (-130 to -60 mV in steps of 10 mV). The current during a voltage step was measured 25 ms after the beginning of the step and was plotted as a function of the voltage. Open symbols represent data recorded in control solution 5 min after breaking the patch, closed symbols are data recorded in the presence of 20 mM TEA and 4 mM 4-AP. The straight lines are linear regressions through both sets of data. The extrapolated reversal potential is at -24 mV. In seven cells the mean reversal potential was at -17 ± 12 mV, that is, near the theoretical E_K in the prevailing conditions (-24 mV). When the potassium concentrations were changed to set E_K at -110 mV ($[\text{K}]_{\text{in}}$ 90 mM and $[\text{K}]_{\text{out}}$ 1.2 mM), the reversal potential of the current suppressible by 4-AP and TEA was again near E_K , indicating that this current was carried primarily by potassium ions. In experiments in which the sodium salts in the pipette were replaced with choline chloride and Tris-ATP, the currents recorded in the absence or in the presence of 4-AP and TEA were the same at each voltage (that is, open and filled circles were superimposed (data not shown)). In seven cells the mean potassium conductance activated by intracellular sodium was 2.4 ± 1.0 nS. Note that this value is lower than the potassium conductance calculated in Fig. 2.

5 mM (or, on occasion, 20 mM). Activation of a potassium current via a drop in ATP^{15} also seems unlikely as the current was observed when recording with pipettes containing 5 mM ATP. Thus, we shall provisionally assume that the potassium current is activated directly by an increase in $[\text{Na}]_{\text{in}}$ and we call it $I_{K(\text{Na})}$. $I_{K(\text{Na})}$ is activated within the first millisecond after the beginning of the sodium current (see Fig. 2), which corresponds to the time required for the sodium current to drive the membrane voltage to the peak of the action potential, as can be seen in current clamp recordings of these neurones. Therefore, our results suggest that during a single action potential, the increase in sodium concentration near the membrane is sufficient to activate $I_{K(\text{Na})}$. Excitatory synaptic events leading to intracellular sodium accumulation (acetylcholine-modulated channels, for example) may also activate $I_{K(\text{Na})}$. As this current is apparently not inactivated even at depolarized potentials, it may help to control membrane repolarization in situations where other potassium currents, such as I_A , are partially or completely inactivated¹. One reason that $I_{K(\text{Na})}$ has not been described in voltage-clamp studies of axons might be that the current is not present in all species or in all neurones. Another possibility is that the channels responsible for it are restricted to the cell soma and possibly to the dendritic tree of neurones.

We thank Drs F. Baumann and A. C. Kato for their comments on the manuscript. This work was supported by a grant from the Swiss NSF (3.651-0.84).

Note added in proof: Since the submission of this paper the effect of sodium influx on a transient outward current has been reported in invertebrate neurones¹⁷.

Received 27 February; accepted 20 August 1985.

1. Connor, J. A. and Stevens, C. F. *J. Physiol., Lond.* **213**, 1-19, 21-30 (1971).
2. Salkoff, L. *Cold Spring Harb. Symp. quant. Biol.* **48**, 222-231 (1983).
3. Schwartzkroin, P. A. & Prince, D. A. *Brain Res.* **185**, 169-181 (1980).
4. Galvan, M., Grafe, P. & Ten Bruggencate, G. T. in *Physiology and Pharmacology of Epileptic Phenomena* (eds Klee, M. R., Lux, H. D. & Speckmann, E. J.) 353-350 (Raven, New York, 1982).
5. Kameyama, M., Kakei, M., Sato, R., Shibasaki, T., Matsuda, H. & Irisawa, H. *Nature* **302**, 354-356 (1984).
6. Bader, C. R., Bertrand, D. & Kato, A. C. *Dev. Biol.* **94**, 131-141 (1982).
7. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1981).
8. Bader, C. R., Bertrand, D., Dupin, E. & Kato, A. C. *Nature* **305**, 808-810 (1983).
9. Fenwick, E. M., Marty, A. & Neher, E. *J. Physiol., Lond.* **331**, 577-597 (1981).
10. Bader, C. R. & Bertrand, D. *J. Physiol., Lond.* **347**, 611-631 (1984).
11. Oxford, G. S., Wu, C. H. & Narahashi, T. *J. gen. Physiol.* **71**, 227-247 (1978).
12. Patlak, J. & Horn, R. J. *J. gen. Physiol.* **79**, 333-351 (1982).
13. Byerly, L., Meech, R. & Moody, W. Jr *J. Physiol., Lond.* **351**, 199-216 (1984).
14. Meech, R. W. A. *Rev. Biophys. Bioengng* **7**, 1-18 (1978).
15. Noma, A. *Nature* **305**, 147-148 (1983).
16. Narahashi, T., Moore, J. W. & Scott, W. R. *J. gen. Physiol.* **47**, 965-974 (1965).
17. Hartung, K. *Pflügers Arch. ges. Physiol.* **404**, 41-44 (1985).

A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16

S. T. Reeders*, M. H. Breuning†, K. E. Davies*, R. D. Nicholls‡, A. P. Jarman‡, D. R. Higgs‡, P. L. Pearson† & D. J. Weatherall*‡

* Nuffield Department of Clinical Medicine, and ‡ MRC Molecular Haematology Unit, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

† Instituut voor Anthropogenetica, Sylvius Laboratoria, Rijksuniversiteit, Leiden, The Netherlands

Adult polycystic kidney disease (APCKD) is a common and often lethal multi-organ disease with an autosomal dominant pattern of inheritance; approximately 1 in 1,000 people carry the mutant gene¹. The major pathological abnormality is the development and progressive enlargement of cysts in several organs including the liver, pancreas and spleen as well as the kidneys. The basic biochemical defect which leads to the formation of cysts remains unknown². Cyst development, which is not retarded by any known therapy, leads to irreversible renal failure and death at a mean age of 51 unless dialysis or transplantation are used³. Patients with the disease account for 9% of chronic dialysis requirement⁴. The first symptoms tend to occur in the fourth decade, after most patients have reproduced³. Presymptomatic diagnosis depends on the ultrasonographic detection of cysts, but exclusion cannot be achieved by this means; 34% of at-risk patients in the second decade and 14% in the third will go on to develop cysts after negative diagnosis⁵. The low sensitivity of diagnostic techniques in this critical age-range imposes severe limitations on genetic counselling and the condition cannot be identified prenatally. Hence we have searched for a linkage marker for APCKD; we show here that the APCKD locus is closely linked to the α -globin locus on the short arm of chromosome 16 ($\hat{\chi}^2 = 25.85$, $\theta = 0.05$).

Four families were identified from the dialysis and transplant registers of the Oxford Renal Unit (NDM-A-D), and five families from units in the Netherlands (NL1-5). These families represent the most typical clinical form of the disease, with the development of end-stage renal failure at a mean age of 48 yr. Diagnosis of symptomatic cases was confirmed by intravenous pyelography or ultrasonography. Asymptomatic at-risk subjects over 15 yr old were assessed by ultrasonography after informed consent had been obtained. Diagnostic criteria were those used by Bear *et al.*⁵, selected to minimize the rate of false-positive diagnosis. Blood samples were obtained from 183 family members. Paternity was confirmed by using mini-satellite probes⁶ in families NDM-A-D.

Our strategy was to use highly polymorphic DNA probes to maximize the efficiency of the search for the APCKD locus⁷. Higgs *et al.*⁸ have previously described a highly polymorphic

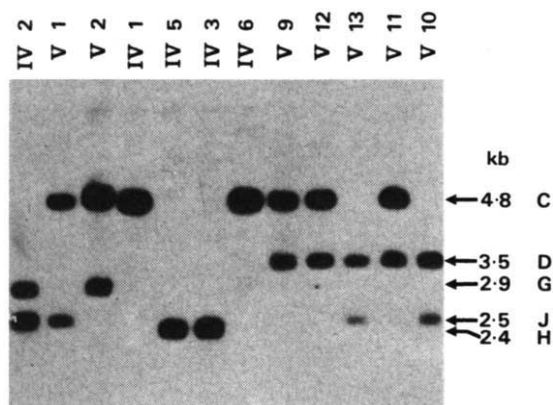


Fig. 1 Autoradiograph of 3'HVR probe hybridized to 5 µg genomic DNA digested with *PvuII*. A 4-kb *HinfI* fragment containing the entire 3'HVR was subcloned into the vector pSP64; the insert was used in hybridization experiments. Members of part of pedigree NDM-A are identified above each track. Alleles distinguishable in this pedigree are named A-L; C, D, G, J and H are shown in Fig. 2. Approximate frequencies for each allele were determined from 52 unrelated Northern Europeans. As the identity of electrophoretically indistinguishable alleles in a highly polymorphic system cannot be assumed, allele frequencies were determined by grouping bands within 0.5- or 1-kb intervals: 1-2 kb, 0.02; 2-2.5 kb, 0.22; 2.5-3 kb, 0.17; 3-4 kb, 0.15; 4-5 kb, 0.14; 5-6 kb, 0.16; 6-7 kb, 0.10; 7-8 kb, 0.04. These estimates, used in lod-score calculation, overestimate the frequency of individual alleles; they therefore reduce the information content of the data and depress the lod score. Methods were as described elsewhere¹⁶ except for the following modifications: DNA was blotted onto GeneScreen Plus (NEN) filters and, after hybridization, filters were washed for 1 h in 2×SSC at 37 °C, and for 1 h in 0.1×SSC at 65 °C.

region ~8 kilobases (kb) beyond the 3' end of the α -globin gene cluster. This region (3' HVR), which has subsequently been cloned and partially sequenced, consists of a tandem-repeat sequence; polymorphism at this locus is attributed to variability in the copy number of the elements of the repeat (R.D.N., A.P.J. and D.R.H., in preparation). Using the Southern blotting technique, a ³²P-labelled *HinfI* fragment including ~4 kb of the 3' HVR was hybridized to *PvuII*-digested genomic DNA. The probe hybridized to two allelic fragments in each individual (Fig. 1). The genomic *PvuII* fragments containing the 3'HVR varied between 1.8 and 8 kb and, in all meioses tested, segregated according to mendelian laws. Genotypes of individuals from family NDM-A are shown in Fig. 2. In this pedigree, 10 electrophoretically distinguishable 3'HVR alleles were seen (A-L), and the disease segregated with allele C. As Fig. 2 shows, it is extremely likely that IV:26 is a recombinant, the disease sub-

Fig. 2 Pedigree of NDM-A. At-risk members identified by empty symbols had negative diagnoses at the age indicated in parentheses below the symbol. The probability that these subjects carry the APCKD gene was assigned as a function of age at diagnosis as follows⁵: age 15-20 yr, 0.34; 20-30 yr, 0.14; 30+yr, 0.05. These estimates were used in the age-of-onset correction option of Liped¹². Family members not at risk were assumed to have an APCKD gene frequency of 0.001 (mean frequency in population)¹. The genotypes at the 3'HVR locus are given for each subject tested. The estimated frequencies for each allele are given in the legend to Fig. 1.

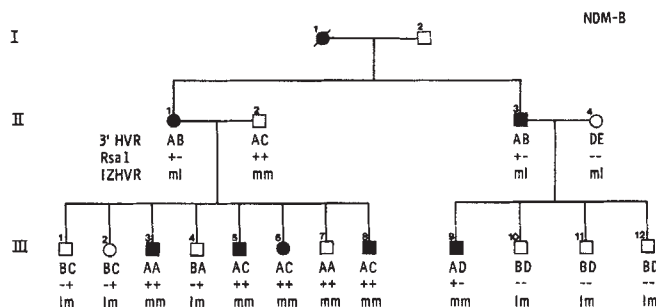
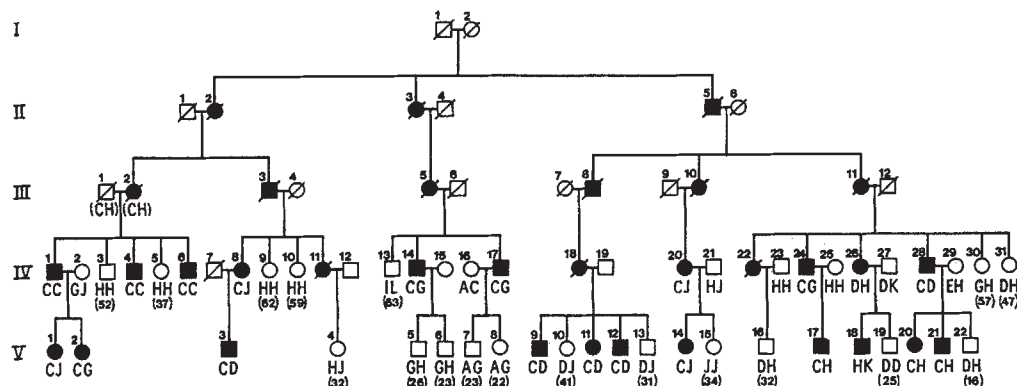


Fig. 3 Members of families NDM-A and -B were informative for two polymorphisms within the α -globin cluster. Part of pedigree NDM-B is shown. The *RsaI* restriction fragment length polymorphism⁹ was detected with a 1.5-kb *PstI* fragment containing the α -globin gene, giving two alleles (+, -). The frequency of the + allele is 0.47 (unpublished data). An inter- ζ length polymorphism was detected in *PvuII*-digested genomic DNA with a *AluI* fragment of the inter- ζ hypervariable region probe (IZHVR). Two alleles were observed (*m*, *l*), with frequencies of 0.68 and 0.32, respectively. These polymorphisms have been described previously^{9,10}. In family NDM-B, five alleles were observed at the 3'HVR locus (A-E), and five haplotypes were obtained (A+m, B-l, C+m, D-m, E-l). APCKD segregates with A+m, except in III:7 who probably represents a crossover between the 3'HVR and APCKD loci.

sequently segregating with allele H in her two sons, V:18 and V:19.

Three-generation families were not available for study because of premature death in this disease, and because we excluded those family members under 15 yr, as reliability of diagnosis is markedly reduced below this age. Meioses were therefore scored by inferring phase. Seven recombinants were observed in 127 meioses (4/66 affected). In nine extended pedigrees, only three meioses were uninformative at the 3'HVR locus, demonstrating the high information content of this polymorphism.

Four additional polymorphic loci within the α -globin cluster were examined in families NDM-A and -B, so as to obtain haplotypes. Two of these sites^{9,10} were polymorphic in members of these families. No crossovers were observed between the 3'HVR and the polymorphic loci in the α -globin cluster. Haplotype data for part of family NDM-B are shown in Fig. 3. In pedigree NDM-A (data not shown), the haplotype C-l was observed in all members who carried the 3'HVR allele C, providing evidence for the identity of this allele in all branches of the family.

The data were analysed for linkage of the 3'HVR marker to the APCKD gene using the computer program LIPED 3 (refs 11, 12). lod scores were calculated using an age-of-onset correction; at-risk individuals who did not meet the criteria for positive

Table 1 lod scores for linkage between 3' HVR and APCKD

	Recombination fraction θ										
	0.00	0.01	0.03	0.04	0.05	0.07	0.10	0.15	0.20	0.30	0.40
Male	$-\infty$	5.33	6.92	7.12	7.23	7.31	7.14	6.60	5.75	3.83	1.66
Female	$-\infty$	15.81	16.40	16.41	16.34	16.03	15.29	13.68	11.80	7.59	3.13
Combined	$-\infty$	23.40	25.60	25.81	25.85	25.56	24.59	22.25	19.48	12.81	5.48

diagnosis were assigned a probability of having inherited the APCKD gene as a function of age⁵, so as to allow for the age-dependent sensitivity of ultrasonography as discussed above and in Fig. 2.

lod scores (Table 1) demonstrate linkage between 3'HVR and the APCKD locus. The maximum lod score for linkage of 3'HVR to APCKD was 7.31 at $\theta = 0.07$ (males) and 16.41 at $\theta = 0.04$ (females). The maximum combined lod score was 25.85 at $\theta = 0.05$ (99% confidence limits 2–11 centimorgans).

Botstein *et al.*⁷ described a theoretical strategy for genetically mapping inherited traits using DNA sequence polymorphisms. Gusella *et al.*¹³ used this approach successfully when they assigned Huntington's disease to chromosome 4. Using similar methods, we have mapped APCKD to within 11 centimorgans of the α -globin cluster, which has previously been assigned to the short arm of chromosome 16^{14,15}. Before applications of the 3'HVR marker can be considered, we must determine whether there is a single locus for all phenotypic variants of the disease. The fact that close linkage was seen in each of the nine families examined suggests a single locus. However, these families represent the commonest phenotype, and further studies are needed, particularly in families with unusual clinical manifestations of the disorder, to assess its genetic heterogeneity.

The localization of the mutation in APCKD to this region of chromosome 16 should facilitate the production of further pro-

bes for presymptomatic detection or prenatal diagnosis of this condition and should ultimately lead to a better understanding of its pathogenesis.

We acknowledge the support of the MRC, Des Oliver and the many physicians who have allowed us to study their patients, and the Netherlands Kidney Foundation. We thank Sarah Ball and Robin Winter for help in analysing the data.

Received 10 July; accepted 19 August 1985.

1. Danovitch, G. M. *Cystic Disease of the Kidney* (ed. Gardner, K. D.) 125–150 (Wiley, New York, 1976).
2. Zerres, K., Völpe, M.-C. & Weiss, H. *Hum. Genet.* **68**, 104–135 (1984).
3. Dalgaard, O. Z. *Acta med. scand.* **158**, Suppl. 238, 1–256 (1957).
4. Lowrie, E. G. & Hampers, C. L. *New Engl. J. Med.* **305**, 434–438 (1981).
5. Bear, J. C. *et al. Am. J. med. Genet.* **18**, 45–53 (1984).
6. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **314**, 67–73; **316**, 76–78 (1985).
7. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. *Am. J. hum. Genet.* **32**, 314–331 (1980).
8. Higgs, D. R., Goodbourn, S. E. Y., Wainscoat, J. S., Clegg, J. B. & Weatherall, D. J. *Nucleic Acids Res.* **9**, 4213–4224 (1981).
9. Higgs, D. R., Hill, A. V. S., Bowden, D. K., Weatherall, D. J. & Clegg, J. B. *Nucleic Acids Res.* **12**, 6965–6977 (1984).
10. Goodbourn, S. E. Y., Higgs, D. R., Clegg, J. B. & Weatherall, D. J. *Molec. Biol. Med.* **2**, 223–238 (1984).
11. Ott, J. *Am. J. hum. Genet.* **26**, 588–597 (1974).
12. Hodge, S. E. *et al. Am. J. hum. Genet.* **31**, 761–762 (1979).
13. Gusella, J. F. *et al. Nature* **306**, 234–238 (1983).
14. Deisseroth, A. *et al. Cell* **12**, 205–218 (1977).
15. Barton, P., Malcolm, S., Murphy, C. & Ferguson-Smith, M. A. *J. molec. Biol.* **156**, 269–278 (1982).
16. Old, J. M. & Higgs, D. R. *Methods in Haematology* (ed. Weatherall, D. J.) 103–113 (Churchill-Livingstone, London, 1983).

Breakpoints in the human T-cell antigen receptor α -chain locus in two T-cell leukaemia patients with chromosomal translocations

William H. Lewis*, Effie E. Michalopoulos*, Dorothy L. Williams†, Mark D. Minden‡ & Tak W. Mak‡

* Departments of Surgery and Microbiology, University of Toronto, Toronto, Ontario, Canada M5S 1A8

† Division of Pathology and Laboratory Medicine, St Jude Children's Research Hospital, Memphis, Tennessee 38101, USA

‡ Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 500 Sherbourne Street, Toronto, Ontario, Canada M4X 1K9

Specific chromosomal translocations have been observed in several human and animal tumours and are believed to be important in tumorigenesis^{1,2}. In many of these translocations the breakpoints lie near cellular homologues of transforming genes, suggesting that tumour development is partly due to the activation of these genes. The best-characterized example of such a translocation occurs in mouse plasmacytoma and human B-cell lymphoma, where *c-myc*, the cellular homologue of the viral oncogene *myc*, is brought into close proximity with either the light- or heavy-chain genes of the immunoglobulin loci^{3–6}, resulting in a change in the regulation of the *myc* gene⁷. T-cell malignancies also have characteristic chromosomal abnormalities, many of which seem to involve the 14q11–14q13 region^{8–12}. This region has recently been found to contain the α -chain genes of the human T-cell antigen receptor^{13–15}. Here we determine more precisely the chromosome break-

points in two patients whose leukaemic T cells contain reciprocal translocations between 11p13 and 14q13. Segregation analysis of somatic cell hybrids demonstrates that in both patients the breakpoints occur between the variable (V) and constant (C) region genes of the T-cell receptor α -chain locus, resulting in the translocation of the C-region gene from chromosome 14 to chromosome 11. As the 11p13 locus has been implicated in the development of Wilms' tumour^{16,17}, it is possible that either the Wilms' tumour gene or a yet unidentified gene in this region is involved in tumorigenesis and is altered as a result of its translocation into the T-cell receptor α -chain locus.

As the α -chain genes of the T-cell receptor have recently been localized to the 14q11–14q13 region, we undertook to determine whether the α -chain locus is involved in such translocations. We identified two patients with T-cell acute lymphoblastic leukaemia (T-ALL) whose leukaemic cells contained a reciprocal translocation involving the short arm of chromosome 11 and the long arm of chromosome 14 near the centromere (Fig. 1)⁸. Somatic cell hybrids were formed between the patients' leukaemic cells and mouse thymidine kinase-negative L cells (Ltk⁻ cells). Independent hybrid clones were screened to detect those containing one of the translocated chromosomes. To do this, we used molecular probes for the catalase gene, which lies in band 11p13 (ref. 18), and the Ha-ras-1 gene, located in the more distal half of 11p (Fig. 1)^{19,20}. Segregation of these two human genes in hybrid clones would indicate the presence of the 11p+ or the 14q- chromosome. Five hybrid clones showed segregation of the catalase and Ha-ras-1 genes (Table 1). Karyotype analysis performed on two of these hybrids confirmed the presence of the 11p+ chromosome in hybrid H11-9b and the 14q- chromosome in hybrid H8-4; neither hybrid contained a normal human chromosome 11 or 14.

To determine whether the breakpoint on chromosome 14 is near the T-cell receptor α -chain locus, DNA from the hybrids

Table 1 Segregation analysis of hybrid clones formed from the fusion of T-ALL cells with mouse cells

Cells	Human 11p marker			T-cell receptor α -chain genes			Chromosomes			
	Catalase	Ha-ras-1	Calcitonin	Constant	Joining	Variable	11	14	11p+	14q-
Parental:										
ALL-8	+	+	+	+	+	+	+	+	+	+
ALL-11	+	+	+	+	+	+	+	+	+	+
Mouse Ltk ⁻	-	-	-	-	-	-	-	-	-	-
Hybrids:										
H11-9b	+	-	-	+	+	-	-	-	+	-
H11-1b	+	-	-	+	+	-	ND	ND	ND	ND
H11-8b	-	+	ND	-	-	+	ND	ND	ND	ND
H8-4	-	+	ND	-	-	+	-	-	-	+
H8-14	-	+	+	-	-	+	ND	ND	ND	ND
H11-1	-	+	+	+	+	+	-	+	-	+
H11-7	+	+	+	-	-	-	+	-	-	-

Leukaemic T cells (86–95% blast cells) were obtained from two patients, ALL-8 and ALL-11, who have been described previously (cases 2 and 3, respectively, in ref. 8). Approximately 4×10^7 cells were fused with an equal number of mouse Ltk⁻ cells using polyethylene glycol 1500; hybrid colonies were selected in growth medium containing hypoxanthine, aminopterin and thymidine^{18,23}. Independent hybrid clones were expanded and recloned in the absence of selective medium. DNA was extracted from 30 recloned hybrids and subjected to Southern analysis to determine the presence of a human-specific 6.6-kb *Bam*HI fragment hybridizing to a Ha-ras-1-specific probe or a 9.0-kb *Pst*I fragment hybridizing to a catalase-specific probe or an 8.0-kb *Taq*I fragment hybridizing to a calcitonin-specific probe¹⁸. Five of the hybrids showed segregation of the human catalase and Ha-ras-1 genes, indicating the presence of either the 11p+ or 14q- chromosomes. For several hybrids the presence of the 11p+ or 14q- chromosomes in the absence of the normal chromosome 11 or 14 was confirmed by detailed karyotyping after trypsin-Giemsa banding. The presence of the T-cell receptor α -chain genes was determined by Southern blot of *Bam*HI-digested genomic DNA followed by hybridization to molecular probes for the C, J and V regions as described in Fig. 2 legend. Hybrids H8-4 and H8-14 were from the fusion with cells of patient ALL-8, while hybrids H11-9b, H11-1b, H11-7, H11-8b and H11-1 were from the fusion with cells from patient ALL-11. ND, not determined.

was examined by Southern blot analysis using three molecular probes specific to the α -chain region. Probes specific for the V and C regions were isolated from DNA complementary to the α -chain T-cell receptor, pY14 (ref. 13). The third probe was a genomic joining (J) region probe: this single copy segment is located 17 kilobases (kb) 5' to the C region of the α -chain²¹ (Table 1, Fig. 2). Genomic DNA digested with the restriction enzyme *Bam*HI and hybridized to the C-region probe showed a single band of 5.5 kb in human DNA (Fig. 2a, lane 1) and one of 7 kb in mouse DNA (Fig. 2a, lane 2). Examination of *Bam*HI-digested DNA from the hybrids revealed that those clones carrying the 11p+ chromosome (H11-9b, lane 4; H11-1b, lane 5) contained the human C-region band in addition to the mouse-specific band, while hybrids containing the 14q- chromosome (H8-14, lane 3; H8-4, lane 6) contained only the mouse 7.0-kb band.

The J-region probe hybridized to a band of 4 kb in *Bam*HI-digested human DNA (Fig. 2b, lane 1). As with the C-region probe, the J-region band was present in hybrids containing the 11p+ chromosome (H11-9b; Fig. 2b, lane 5) but absent from those hybrids with the 14q- chromosome (H8-14; Fig. 2b, lane 3).

When the V-region probe was used, a reciprocal result was obtained. Figure 2c illustrates that while with *Bam*HI 15- and 4-kb bands characteristic of human DNA (lanes 1, 2) were present in hybrids carrying the 14q- chromosome (H8-14, lane 4; H8-4, lane 7), they were absent from the hybrids containing the 11p+ chromosome (H11-9b, lane 5; H11-1b, lane 6).

In all hybrids tested, the presence of the 11p+ chromosome was correlated with the presence of the T-cell receptor α -chain constant and J-region sequences (Table 1). Similarly, the presence of the 14q- chromosome, in the absence of a normal chromosome 14, was correlated with the presence of the T-cell receptor α -chain V-region genes. Taken together, the data presented here indicate clearly that the translocation of chromosome 11 with chromosome 14 in both T-ALL patients separates the V and C regions of the T-cell receptor α -chain, thus localizing the breakpoints within the T-cell receptor α -chain locus. The segregation of the genomic J region sequences with the C region indicates that the breakpoints occur more than 17 kb 5' of the C region²¹.

The above results allow us to derive the orientation of the T-cell receptor α -chain locus on chromosome 14. The proposed

order is: centromere; V region; J region; C region; telomere (Fig. 1).

The regional localization of the breakpoints on chromosome 14 for the leukaemic cells used in the present study have been tentatively assigned⁸ previously to band 14q13. However, several *in situ* hybridization studies^{14,15} have subsequently located the T-cell receptor α -chain gene closer to the centromere in band 14q11. This discrepancy may be a result of the small size of the derivative 14q- chromosome, which led to an imprecise assignment of the translocation breakpoint. The data presented here, however, clearly demonstrate, independently of band assignment, that the translocation associated with T-cell ALL separates the T-cell receptor α -chain C region from the V region.

The exact gene of chromosome 11 that may be affected by this translocation is not known. The results indicate that for both patients the breakpoint on chromosome 11 occurs distal to the catalase gene. Preliminary data indicate that the breakpoint occurs between the catalase and the calcitonin genes (Table 1). As calcitonin has been localized¹⁸ to band 11p14, this places the T-ALL breakpoint close to the putative Wilms' tumour locus, which it has been suggested²² lies distal to catalase within 11p13. Further studies are necessary to determine whether the

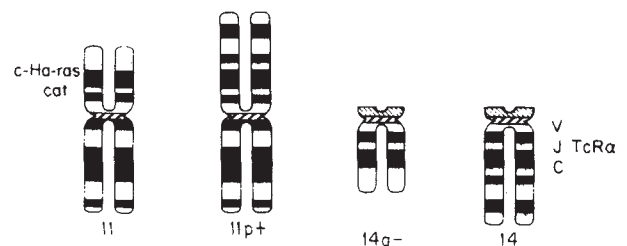


Fig. 1 Schematic representation of the chromosome 11;14 reciprocal translocation associated with T-cell ALL. The detailed karyotypes of the leukaemic cells from the patients used in the present study have been described previously⁸. Letters on the left indicate the approximate locations of the Ha-ras-1 gene (c-Ha-ras) and the catalase gene (cat) on the normal chromosome 11. The letters on the right indicate the order of the α -chain T-cell antigen receptor (TcRa) genes for the variable (V), joining (J) and constant (C) regions on the normal chromosome 14.

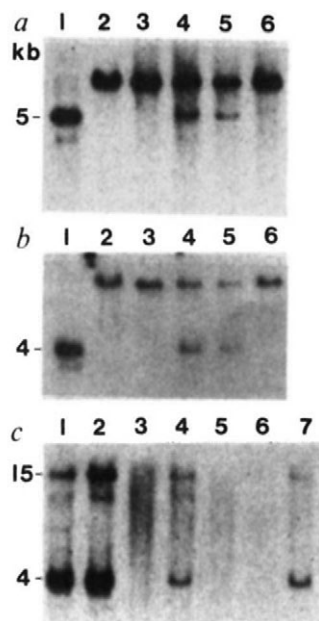


Fig. 2 Autoradiograms of Southern blots of genomic DNA digested with *Bam*HI and hybridized to molecular probes for the T-cell receptor α -chain genes for the C region (a), J region (b) and V region (c). The C-region probe consisted of a 600-base pair (bp) *Pvu*II fragment of the T-cell receptor α -chain cDNA pY14, and the V-region probe consisted of a 400-bp *Pvu*II fragment of the same cDNA. The J-region probe was a 2.3-kb *Eco*RI fragment derived from a genomic region 17 kb 5' to the constant region. Hybridizations and filter washings were carried out as described elsewhere¹⁸ except that the final wash for the C-region and J-region hybridizations was 0.6 $\times$ SSC for 30 min at 65°C and for the V-region hybridization the final wash was 0.3 $\times$ SSC for 30 min at 65°C. Longer exposures of filters hybridized to the V-region probe revealed characteristic mouse bands in Ltk⁻, H11-9b and H11-1b. a, Lane 1, human DNA; lane 2, mouse DNA; lane 3, hybrid H8-14; lane 4, H11-9b; lane 5, H11-1b; lane 6, H8-4. b, lane 1, human; lane 2, mouse; lane 3, hybrid H8-14; lane 4, H11-1; lane 5, H11-9b; lane 6, H11-7. c, Lane 1, T-cell DNA from patient ALL-8; lane 2, T-cell DNA from patient ALL-11; lane 3, mouse; lane 4, hybrid H8-14; lane 5, H11-9b; lane 6, H11-1b; lane 7, H8-4.

same gene on chromosome 11 involved in T-ALL is involved in Wilms' tumour. In either case, the translocation of chromosome 11 sequences into the T-cell α -chain locus provides an opportunity to isolate the gene.

This work was supported by grants from the NCI of Canada and the MRC of Canada. We thank Dr H. Willard and V. Powers for performing the karyotyping and Drs Y. Yoshikai and N. Caccia for providing the molecular probes. W.H.L. thanks Drs Bernie Langer, Judy Falk and Rudy Falk of the Department of Surgery, University of Toronto, for their continued support. M.D.M. is a Scholar of the Leukemia Society of America.

Received 7 May; accepted 12 August 1985.

- Yunis, J. J. *Science* **221**, 227-235 (1983).
- Klein, G. *Cell* **32**, 311-315 (1983).
- Cory, S., Gerondakis, S. & Adams, J. M. *EMBO J.* **2**, 697-703 (1983).
- Marcu, K. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 519-523 (1983).
- Croce, C. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 6922-6926 (1983).
- Hollis, G. F. *et al. Nature* **307**, 752-754 (1984).
- Taub, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 7837-7841 (1982).
- Williams, D. L. *et al. Cell* **36**, 101-109 (1984).
- Clare, N. C., Boldt, D. H., Messerschmidt, G., Hansen, K. L. & Manhoff, L. *Lab. Invest.* **52**, 14A (1985).
- McCaw, B., Hecht, F., Harnden, D. G. & Teplitz, R. L. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2071-2075 (1975).
- Zech, L. *et al. Nature* **308**, 858-860 (1984).
- Ueshima, Y., Rowley, J. D., Variakojis, D., Winter, J. & Gordon, L. *Blood* **63**, 1028-1038 (1984).
- Yanagi, Y., Chan, A., Chin, R., Minden, M. D. & Mak, T. W. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3430-3434 (1985).

- Caccia, N. *et al. J. exp. Med.* **161**, 1255-1260 (1985).
- Croce, C. *et al. Science* **227**, 1044-1047 (1985).
- Francke, U., Holmes, L. B., Atkins, L. & Riccardi, V. M. *Cytogenet. cell. Genet.* **24**, 185-192 (1979).
- Riccardi, V. M. *et al. Cancer Genet. Cytogenet.* **2**, 131-137 (1980).
- Michalopoulos, E. E. *et al. Hum. Genet.* **70**, 157-162 (1985).
- de Martinville, B. & Francke, U. *Nature* **305**, 641-643 (1983).
- Huerre, C., Despoisse, S., Gilgenkratz, S., Lenoir, G. M. & Junien, C. *Nature* **305**, 638-641 (1983).
- Yoshikai, Y. *et al. Nature* **316**, 837-840 (1985).
- Turlau, C. *et al. Hum. Genet.* **67**, 455-456 (1984).
- Falk, J. A., Lewis, W. H. & Letarte, M. *Hum. Immun.* **13**, 265-280 (1985).

A model for intracellular translocation of protein kinase C involving synergism between Ca^{2+} and phorbol esters

Marlene Wolf, Harry LeVine III, W. Stratford May Jr*, Pedro Cuatrecasas & Naji Sahyoun

Department of Molecular Biology, Wellcome Research Laboratories, Burroughs Wellcome Co., Research Triangle Park, North Carolina 27709, USA

* Johns Hopkins Oncology Center, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA

The activation of protein kinase C by diacylglycerol and by tumour promoters has implicated this enzyme in transmembrane signalling and in the regulation of the cell cycle^{1,2}. *In vitro* studies revealed that catalytic activity requires the presence of calcium and phospholipids with a preference for phosphatidylserine³. Diacylglycerol and tumour promoters such as phorbol esters bind to the enzyme⁴⁻⁷, leading to its activation while sharply increasing its affinity for Ca^{2+} and phospholipid. Addition of diacylglycerol analogues or phorbol esters to intact cells results in the phosphorylation of specific polypeptides^{4,8-20}. Several cellular processes, including hormone and neurotransmitter release and receptor down-regulation¹⁻⁴, are modulated by the activation of protein kinase C, while phorbol ester-induced stimulation of the enzyme in whole cells has been associated with its translocation from the cytoplasm to the plasma membrane^{21,22}. Moreover, the use of Ca^{2+} ionophores has revealed an apparent synergism between Ca^{2+} mobilization and protein kinase C activation². This synergism has recently also been found to apply to receptor down-regulation (ref. 23 and accompanying paper²⁴). Here we describe a reconstitution system in which intracellular translocation of protein kinase C and the synergism between Ca^{2+} and enzyme activators can be studied. The results suggest a rationale for concomitant Ca^{2+} mobilization and diacylglycerol formation in response to some hormones, neurotransmitters and growth factors.

Inside-out human erythrocyte vesicles are useful for studying physiologically relevant protein-protein and protein-lipid interactions. Furthermore, these erythrocyte membrane vesicles lack endogenous phorbol ester binding activity, rendering them particularly suitable for reconstitution with exogenous phorbol ester receptors. Accordingly, binding of purified protein kinase C to inside-out vesicles from human erythrocytes was considered as a model for intracellular translocation of the enzyme. The amount of enzyme bound was conveniently determined by its ability to bind ³H-4- β -phorbol-12,13-dibutyrate with an exact 1:1 stoichiometry⁶. The binding of the enzyme to the membranes occurs at relatively high concentrations of Ca^{2+} (10-100 μM) in the absence of Mg^{2+} (Fig. 1a). However, the presence of 3 mM MgCl_2 dramatically increases the efficacy of Ca^{2+} , with protein kinase C binding occurring consistently at Ca^{2+} concentrations between 50 nM and 5 μM (Fig. 1b). The sharpest increase in the binding reaction occurs between 100 nM and 500 nM free Ca^{2+} (Fig. 1b, c), agreeing favourably with the reported levels of intracellular Ca^{2+} in basal and various stimulated conditions, respectively. Similar observations were made

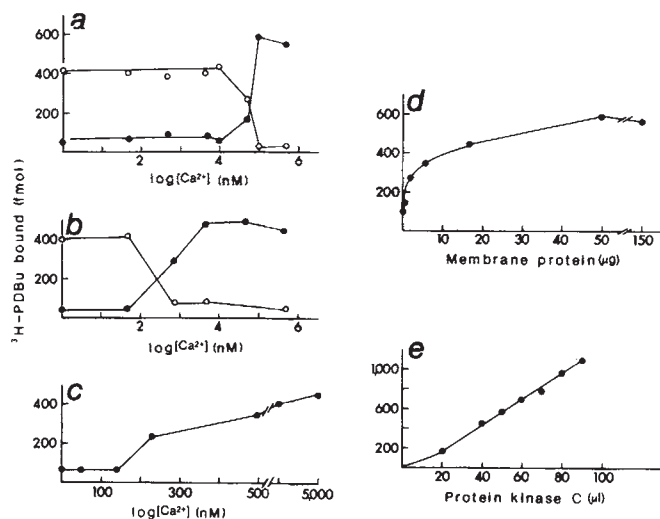


Fig. 1 Calcium-dependent binding of protein kinase C to red cell membranes. *a*, Binding of protein kinase C to the membranes in the absence of MgCl_2 . ●, Bound enzyme; ○, free enzyme. *b*, Binding in the presence of 3 mM MgCl_2 . Symbols as in *a*. *c*, Binding of protein kinase C to membranes at low Ca^{2+} concentrations. *d*, Dependence of the binding reaction on membrane concentration using 50 μl protein kinase C. *e*, Dependence of the binding reaction on enzyme concentration using 10 μg of membrane protein.

Methods. Protein kinase C was purified from rat brains as described earlier²⁸. The experiments were performed with the enzyme, eluted from the Mono Q anion exchange column (Pharmacia), which has a specific activity of ~40 pmol $^3\text{H-4-}\beta$ -phorbol-12,13-dibutyrate (PDBu) bound per ml of enzyme, and ~15 nmol ^{32}P incorporated into histone H1 per min per ml; the protein concentration was ~0.3 mg ml^{-1} . Further purification by phenyl-Sepharose chromatography²⁸ did not alter the results. Inside-out vesicles from human erythrocytes were prepared by incubating washed ghosts in 1 mM Tris-HCl pH 7.6, 0.05 mM EDTA at 4°C for 15 h. The inside-out vesicles were washed three times in 10 mM Tris-HCl pH 7.6, and resuspended in this buffer at a concentration of 3 mg ml^{-1} protein. Binding of protein kinase C to the membranes was examined in 1.5-ml Eppendorf tubes in a volume of 200 μl of a buffer containing 10 mM Tris-HCl pH 7.6, 1 mM dithiothreitol, 3 mM MgCl_2 , 30 $\mu\text{g ml}^{-1}$ bovine serum albumin, 100 μM EGTA, 150–300 μg membrane protein and 50 μl protein kinase C; CaCl_2 was added to give the indicated free calcium concentrations²⁹. Zero Ca^{2+} corresponds to 100 μM EGTA. The suspension was incubated at 25°C for 10 min and then centrifuged for 15 min at 40,000g. The supernatant was carefully removed and the membranes were resuspended in 170 μl 10 mM Tris-HCl pH 7.6. The amount of protein kinase C bound to the membranes or remaining in the supernatant was measured by [^3H]-PDBu binding. 50 μl of the resuspended membranes or the supernatant, respectively, was added to 50 μl of a reaction mixture containing 10 mM Tris-HCl pH 7.6, 2 mM dithiothreitol, 10 mM MgCl_2 , 400 μM CaCl_2 , 100 $\mu\text{g ml}^{-1}$ bovine serum albumin, and 60 nM [$^{20-3}\text{H(N)}$]-PDBu (6.5 Ci mol^{-1} , NEN). The reaction mixture for the supernatant also contained 40 $\mu\text{g ml}^{-1}$ phosphatidylserine. The binding was performed for 40 min at 25°C. Bound $^3\text{H-PDBu}$ was separated from free $^3\text{H-PDBu}$ by suction through Whatman GF/C glass filters. The determinations were run in triplicate and the experiments were repeated at least twice, yielding a s.e.m. of <7%.

on lysed synaptosomes and HL-60 leukaemia cell membranes instead of erythrocyte vesicles. These results imply that physiological increases of intracellular Ca^{2+} cause a translocation of protein kinase C from a soluble to a membrane-bound compartment. The subcellular compartment to which the enzyme binds may depend on the precise site of Ca^{2+} mobilization and the resulting local intracellular Ca^{2+} levels. Therefore, stimuli which raise intracellular Ca^{2+} levels can lead to membrane binding of protein kinase C, thus exposing it to inner leaflet phospholipids and to the potential generation of diacylglycerol. This may be

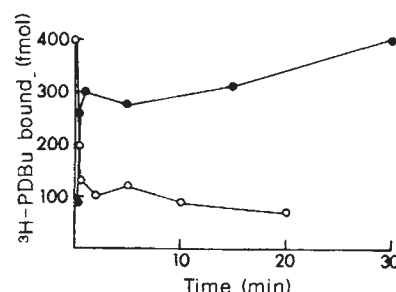


Fig. 2 Time course of association and dissociation of protein kinase C to and from membranes. The binding of protein kinase C to red cell inside-out vesicles was performed as described in Fig. 1 legend with a free Ca^{2+} concentration of 1 μM . At the indicated time, the tubes were placed in a Beckman Microfuge and centrifuged for 30 s. The pelleted membranes were resuspended and assayed for $^3\text{H-PDBu}$ binding (●). The dissociation of membrane-bound protein kinase C was assessed following pre-binding of the enzyme to the membranes in the presence of 5 μM Ca^{2+} . After 10 min at 25°C the membranes were centrifuged and the pellets were resuspended in 200 μl of buffer containing 10 mM Tris-HCl pH 7.6, 3 mM MgCl_2 , 1 mM dithiothreitol, 30 $\mu\text{g ml}^{-1}$ bovine serum albumin and 100 μM EGTA. At the indicated time, the tubes were centrifuged and $^3\text{H-PDBu}$ binding to the resuspended membranes was assayed (○).

viewed as a 'priming' of the protein kinase C system, allowing it to participate in transmembrane signalling.

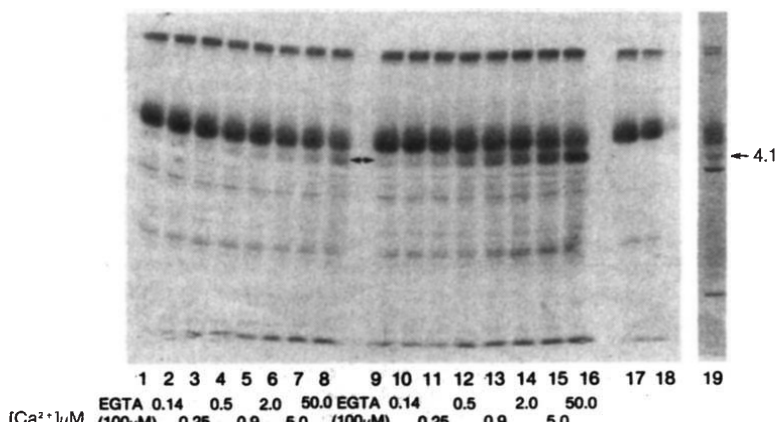
The binding reaction seems to be a linear function of enzyme concentration and of membrane concentration under conditions of excess enzyme, but not excess membranes (Fig. 1*d, e*). At 500 nM Ca^{2+} , enzyme binding is not affected by treating the inside-out vesicles with 50 $\mu\text{g ml}^{-1}$ trypsin, 1 M NaOH or by heating at 90°C. These observations suggest that membrane proteins are not essential for the binding of protein kinase C, although they may serve a regulatory function. Liposomes prepared from human erythrocyte lipids also bind the enzyme, but require Ca^{2+} concentrations greater than 10 μM , implying that disturbing the original membrane bilayer organization results in a markedly reduced sensitivity to Ca^{2+} . These observations favour the use of plasma membrane preparations, rather than derivative liposomes, in the study of protein kinase C-membrane interactions. The apparent dissociation constant of $^3\text{H-4-}\beta$ -phorbol-12,13-dibutyrate was 20 nM for membrane-bound protein kinase C, and a similar value was obtained for the soluble enzyme in the presence of phosphatidylserine (data not shown).

Fast membrane association and dissociation rates of protein kinase C would permit rapid responsiveness to changing intracellular Ca^{2+} levels. In fact, at 1 μM Ca^{2+} , the binding reaction is almost complete within 30 s (Fig. 2). Similarly, when free Ca^{2+} levels are reduced in the presence of excess EGTA, protein kinase C is released mostly during the initial 30 s (Fig. 2).

For the Ca^{2+} -induced membrane binding of protein kinase C to be physiologically relevant the enzyme must be able to interact with endogenous substrates. We therefore examined protein kinase C-dependent phosphorylation of erythrocyte membrane substrates. Figure 3 clearly demonstrates that protein kinase C binding leads to Ca^{2+} -dependent phosphorylation of several polypeptides, with band 4.1 as a preferred substrate; these results are consistent with the phorbol ester-stimulated phosphorylation of band 4.1 in intact rabbit erythrocytes²⁵. The reconstitution system can be adapted to other membrane preparations containing potential protein kinase C substrates. The experimental procedure involves release of endogenous pre-bound enzyme by EGTA, followed by reconstitution with exogenous protein kinase C in the presence of Ca^{2+} .

Interestingly, Ca^{2+} seems to have a dual effect on protein kinase C; membrane binding is maximal at 500 nM Ca^{2+} while enzyme activation occurs mostly between 5 μM and 50 μM Ca^{2+}

Fig. 3 Calcium and phorbol ester dependence of the phosphorylation of membrane proteins. The phosphorylation experiment was performed under conditions similar to those used for the binding experiments, with 200 μ l of 10 mM Tris-HCl pH 7.6, 1 mM dithiothreitol, 3 mM MgCl_2 , 30 $\mu\text{g ml}^{-1}$ bovine serum albumin, 10 μM ATP, 1 μCi (γ - ^{32}P)ATP (50 Ci mmol^{-1} , NEN), 150 μg membrane protein and 50 μl protein kinase C. Calcium concentrations are indicated below the autoradiogram. Lanes 1–8, without PDBu; lanes 9–16, with 100 nM PDBu; lane 17, endogenous phosphorylation with 100 μM EGTA and in the absence of protein kinase C; lane 18, endogenous phosphorylation with 100 μM free Ca^{2+} and 100 nM PDBu in the absence of protein kinase C; lane 19, Coomassie blue-stained gel of the membrane proteins. The reaction was performed at 25 $^\circ\text{C}$ for 10 min and stopped by adding 100 μl SDS buffer, and analysed by SDS-polyacrylamide gel electrophoresis using 7.5% acrylamide and 0.02% bisacrylamide³⁰. Similar results were obtained when protein kinase C was pre-bound to the membranes and the free enzyme removed before phosphorylation. The arrow designates band 4.1. [Ca^{2+}] corresponds to free Ca^{2+} , and EGTA (lanes 1, 9) indicates 100 μM EGTA in the absence of Ca^{2+} .



in the absence of diacylglycerol or phorbol esters (Fig. 3). It is not known whether this discrepancy between the Ca^{2+} -dependent enzyme binding to membranes and enzyme activation by Ca^{2+} arises from the existence of multiple Ca^{2+} binding sites on the enzyme. At intracellular Ca^{2+} levels, the Ca^{2+} -dependent membrane binding reaction is likely to be the more relevant process. Thus, Ca^{2+} -induced protein kinase C binding to membranes may not be a sufficient condition for enzyme activation.

The synergism of intracellular Ca^{2+} mobilization and protein kinase C activation is usually attributed to the requirement for Ca^{2+} -calmodulin, as well as protein kinase C-mediated phosphorylation². This interpretation was originally devised for platelet activation involving phosphorylation of a polypeptide of relative molecular mass (M_r) 40,000 (40K) by protein kinase C and of myosin light chain (M_r 20K) by Ca^{2+} -calmodulin-dependent myosin light-chain kinase. The Ca^{2+} ionophore A23187 apparently did not potentiate the phosphorylation of the 40K- M_r polypeptide by protein kinase C in the presence of phorbol esters or diacylglycerol analogues^{4,8,26}. However, subsequent investigation revealed that Ca^{2+} ionophores do synergize protein phosphorylation by protein kinase C in whole cells. These polypeptides include a 50K- M_r neutrophil protein kinase C substrate²⁷ and the transferrin receptor of HL-60 cells^{23,24}. Synergism of protein phosphorylation closely matched the synergism of neutrophil lysosomal enzyme release²⁷ and transferrin receptor down-regulation^{23,24}, respectively. These findings suggest that elevated intracellular Ca^{2+} levels have a synergistic effect on protein kinase C activation by acting on the enzyme itself.

Results from the reconstitution system are consistent with this suggestion. Protein kinase C binding to erythrocyte inside-out vesicles was studied in the presence or absence of 4- β -phorbol-12,13-dibutyrate. The phorbol ester clearly enhances the sensitivity of the binding reaction to Ca^{2+} at lower Ca^{2+} concentrations, and increases total enzyme binding at higher Ca^{2+} concentrations (Fig. 4a). Moreover, raising the Ca^{2+} concentration from 100 nM to 1.0 μM augments protein kinase C binding at all concentrations of phorbol ester tested (Fig. 4b). Accordingly, Ca^{2+} and phorbol esters seem to act synergistically to effect protein kinase C binding to membranes, a step required for intracellular enzyme activation. The site of endogenous diacylglycerol generation as well as that of Ca^{2+} mobilization may confer some compartmentalization on the synergistic response. However, note that the rate of membrane association of protein kinase C in the presence of the phorbol ester and low (100 nM) Ca^{2+} levels is much slower than the rate achieved at 1 μM Ca^{2+} (Figs 2, 4c). Thus, raising intracellular Ca^{2+} levels will permit a much faster response to external stimuli that

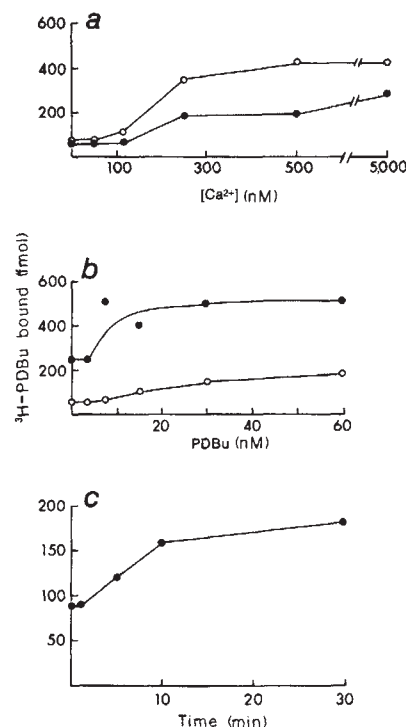
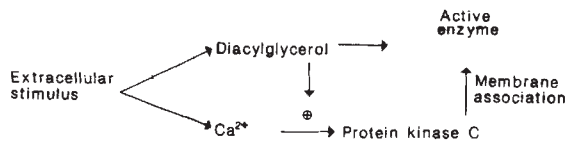


Fig. 4 Calcium and phorbol ester act synergistically on the binding of protein kinase C to membranes. **a**, Effect of PDBu on the Ca^{2+} dependence of protein kinase C binding to membranes. The membrane-binding assay was performed in the presence (○) or absence (●) of 100 nM $^3\text{H-PDBu}$. **b**, Effect of Ca^{2+} on the $^3\text{H-PDBu}$ dependence of protein kinase C binding to membranes. The membrane-binding assay was performed in the presence of 100 nM (○) or 1 μM (●) Ca^{2+} . **c**, Time course of $^3\text{H-PDBu}$ -dependent protein kinase C binding to membranes. Binding was performed in the presence of 100 nM Ca^{2+} and 100 nM $^3\text{H-PDBu}$. The binding of protein kinase C to inside-out vesicles was performed at various Ca^{2+} and PDBu concentrations. When present, PDBu was used in the radioactive form ($^3\text{H-PDBu}$) so as to avoid interference of unlabelled PDBu with the second stage of the experiment in which $^3\text{H-PDBu}$ was used to detect the membrane-bound enzyme.

activate protein kinase C, providing another modality for synergism between intracellular Ca^{2+} and protein kinase C activators. Such a modality may complement previously proposed synergistic pathways between protein kinase C and other Ca^{2+} -activated effector systems².

Conclusions derived from the present results are summarized below, highlighting the concerted action of Ca^{2+} and diacylglycerol or its analogues, leading to protein kinase C activation.



This molecular mechanism may provide another regulatory pathway that can be modulated by intracellular Ca^{2+} .

Received 28 May; accepted 22 August 1985.

1. Takai, Y. *et al.* *Adv. Cyclic Nucleotide Res.* **14**, 301-313 (1981).
2. Nishizuka, Y. *Nature* **308**, 693-698 (1984); *Trends biochem. Sci.* **8**, 13-16 (1983); *Science* **225**, 1365-1370 (1984).
3. Takai, Y. *et al.* *J. biol. Chem.* **254**, 3692-3695 (1979).
4. Castagna, M. *et al.* *J. biol. Chem.* **257**, 7847-7851 (1982).
5. Nield, J. E., Kuhn, L. J. & Vandenbark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **80**, 36-40 (1983).
6. Kikkawa, U., Takai, Y., Tanaka, Y., Miyake, R. & Nishizuka, Y. *J. biol. Chem.* **258**, 11442-11445 (1983).
7. Takai, Y., Kishimoto, A., Kikkawa, U., Mori, T. & Nishizuka, Y. *Biochem. biophys. Res. Commun.* **91**, 1218-1224 (1979).
8. Kaibuchi, K. *et al.* *J. biol. Chem.* **258**, 6701-6704 (1983).
9. Imaoka, T., Lynham, J. A. & Haslam, R. J. *J. biol. Chem.* **258**, 11404-11414 (1983).
10. Trevillyan, J. M., Kulkarni, R. K. & Byus, C. V. *J. biol. Chem.* **259**, 897-902 (1984).
11. Kelleher, D. J., Pessin, J. E., Ruoho, R. E. & Johnson, G. L. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4316-4320 (1984).
12. Jacobs, S., Sahyoun, N. E., Saliel, A. R. & Cuatrecasas, P. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6211-6213 (1983).
13. Iwashita, S. & Fox, C. F. *J. biol. Chem.* **259**, 2559-2567 (1984).
14. Cochet, C., Gill, G. N., Meisenhelder, J., Cooper, J. A. & Hunter, T. *J. biol. Chem.* **259**, 2553-2558 (1984).
15. May, W. S. Jr, Sahyoun, N. E., LeVine, H. III, Jacobs, S. & Cuatrecasas, P. *Clin. Res.* **32**, 498A (1984).
16. May, S., Jacobs, S. & Cuatrecasas, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2016-2020 (1984).
17. Naka, M., Nishikawa, M., Adelstein, R. S. & Hidaka, H. *Nature* **306**, 490-492 (1983).
18. Huang, C.-K., Hill, J. M. Jr, Bormann, B. J., Mackin, W. M. & Becker, E. L. *J. biol. Chem.* **259**, 1386-1389 (1984).
19. Werth, D. K., Nield, J. E. & Pastan, I. *J. biol. Chem.* **258**, 11423-11426 (1983).
20. Kawamoto, S. & Hidaka, H. *Biochem. biophys. Res. Commun.* **118**, 736-742 (1984).
21. Kraft, A. S., Anderson, W. B., Cooper, H. L. & Sando, J. J. *J. biol. Chem.* **257**, 13193-13196 (1982).
22. Kraft, A. S. & Anderson, W. B. *Nature* **301**, 621-623 (1983).
23. May, W. S., Sahyoun, N., Sensenbrenner, L. L. & Cuatrecasas, P. *Blood* **64**, Suppl. 1, 479 (1984).
24. May, W. S., Sahyoun, N., Wolf, M. & Cuatrecasas, P. *Nature* **317**, 549-551 (1985).
25. Ling, E. & Saperstein, V. *Biochem. biophys. Res. Commun.* **120**, 291-298 (1984).
26. Kajikawa, N. *et al.* *Biochem. biophys. Res. Commun.* **116**, 743-750 (1983).
27. White, J. R. *et al.* *J. biol. Chem.* **259**, 8605-8611 (1984).
28. Wolf, M., Sahyoun, N., LeVine, H. III & Cuatrecasas, P. *Biochem. biophys. Res. Commun.* **122**, 1268-1275 (1984).
29. Fabiato, A. & Fabiato, F. *J. Physiol., Paris* **75**, 463-506 (1979).
30. Laemmli, U. K. *Nature* **227**, 680-685 (1970).

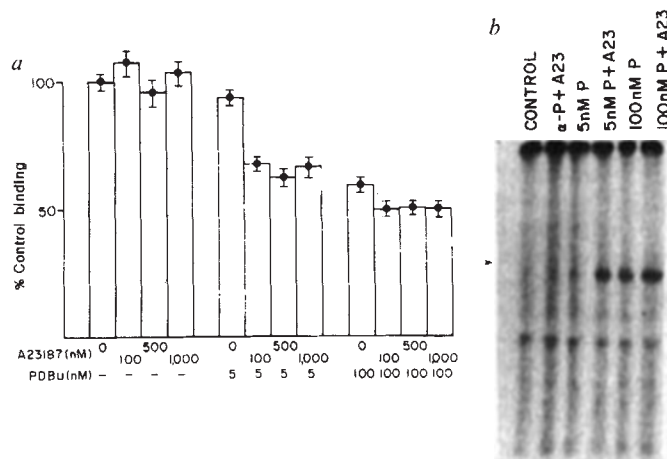


Fig. 1 Effect of A23187 on phorbol dibutyrate-induced down-regulation (a) and phosphorylation (b) of the surface transferrin receptor. Additions to each lane in b were made as follows: Control, no additions; α -P + A23, 100 nM 4- α -phorbol-12,13-dibutyrate to 200 nM A23187; 5 nM P, 5 nM 4- β -phorbol-12,13-dibutyrate; 5 nM P + A23, 5 nM P to 200 nM A23187; 100 nM P, 100 nM P; 100 nM P + A23, 100 nM P to 200 nM A23187. Arrowhead indicates migration of the transferrin receptor of relative molecular mass 92,000.

Methods. a, Whole cells in logarithmic growth were washed once in RPMI 1640 medium containing 10 mM HEPES, pH 7.3 and incubated in the same medium containing 2% fetal calf serum and 1 mM CaNO_3 at 37 °C. Various concentrations of A23187 were added for 5 min before addition of either 5 or 100 nM 4- β -phorbol-12,13-dibutyrate (P) for 60 min. Cells were centrifuged and pellets resuspended into 0.5 ml 50 mM Tris-HCl, pH 7.3, containing 150 mM NaCl and 0.5 mg ml⁻¹ ovalbumin at 4 °C. ¹²⁵I-ferrotransferrin at a concentration of 21 nM was added and cells were incubated for 60 min at 4 °C as described previously⁹. Following incubation, cell pellets were collected by centrifugation for 60 s in a Beckman microfuge. Supernatant containing unbound transferrin was aspirated and the cell pellet counted in a γ counter. Results are expressed as per cent mean control binding $\pm$ s.e.m. Control binding is arbitrarily set at 100% and represents that concentration of ¹²⁵I-ferrotransferrin which binds to cells incubated in the absence of any additions. Total specific binding of ¹²⁵I-ferrotransferrin to control cells was 264 fmol per 10⁶ cells. Nonspecific binding was determined in the presence of 1,000-fold excess unlabelled ferrotransferrin. b, Cells were collected and washed three times in buffer containing 50 mM Tris-HCl, pH 7.3, and 150 mM NaCl and resuspended at a concentration of 5×10^7 cells ml⁻¹ containing 20 mM HEPES, pH 7.3, penicillin and streptomycin. ³²P-orthophosphoric acid (carrier free, Atomic Energy of Canada) was added at 1 mCi ml⁻¹ cells and incubation continued for 2.5 h at 37 °C. Addition of 200 nM A23187 (A23) ionophore was made for 5 min followed by either 5 to 100 nM 4- β -phorbol-12,13-dibutyrate (P) or 200 nM 4- α -phorbol-12,13-dibutyrate (α -P) for 45 min. Cells were washed four times in ice-cold buffer containing 50 mM Tris-HCl, pH 7.3, 150 mM NaCl and 1.5 mM MgCl_2 (wash buffer). The cell pellet was vigorously resuspended in lysis buffer which contained in wash buffer 1.5% Triton X-100, 1 mM phenylmethylsulphonyl fluoride and 1 mg ml⁻¹ bacitracin at 4 °C. Incubation with frequent vigorous vortexing was carried out for 30 min at 4 °C. The solubilized cellular material was separated by centrifugation at 3,000 r.p.m. in a Sorvall RC3B for 20 min. The supernatant was diluted with an equal volume of lysis buffer without Triton X-100. To 1 ml of this supernatant was added 100 μ l of Sepharose-4B affinity gel containing covalently coupled ferrotransferrin as described previously⁹. Incubation was continued for 3 h at 4 °C. The affinity gel was washed five times in buffer containing 0.2% Triton X-100 and the resulting supernatant of each wash was discarded. The pellet was incubated with 250 μ l electrophoresis sample buffer containing 125 mM Tris-HCl, pH 6.8, 20% glycerol, 3% SDS, 200 mM dithiothreitol and 0.05% bromophenol blue. The mixture was incubated at room temperature for 15 min and boiled for 5 min. The supernatant, containing solubilized transferrin receptor, was electrophoresed according to Laemmli¹². The gel was dried and exposed to Kodak X-Omatic film for an optimal period and the autoradiogram is shown here.

synergism between Ca^{2+} and phorbol esters that leads to transferrin receptor phosphorylation and down-regulation in HL-60 human leukaemic cells. Raising intracellular Ca^{2+} , although ineffective by itself, increases the potency and rate of action of phorbol ester for activating protein kinase C and mediating transferrin receptor phosphorylation and down-regulation. We propose a molecular model in which increased intracellular Ca^{2+} recruits protein kinase C to the plasma membrane, thus 'priming' the system for activation by phorbol ester.

The transferrin receptor, a surface glycoprotein, is a specific marker for rapidly proliferating cells²⁰. Phorbol esters not only

Role of intracellular calcium mobilization in the regulation of protein kinase C-mediated membrane processes

W. S. May Jr, N. Sahyoun*, M. Wolf* & P. Cuatrecasas*

Johns Hopkins Oncology Center, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA

* Department of Molecular Biology, Wellcome Research Laboratories, Burroughs-Wellcome Company, Research Triangle Park, North Carolina 27709, USA

Phorbol esters are potent tumour-promoting agents that exert pleiotropic effects on cells¹⁻³. Among these are the control of growth⁴, stimulation of release of stored bioactive constituents⁵⁻⁸ and regulation of growth-factor surface receptors⁹⁻¹³. Phorbol esters bind to and activate protein kinase C, leading to the phosphorylation of specific protein substrates presumed to be necessary for eliciting the full response^{6,8-12,14,15}. Strong evidence exists that specific binding of tumour promoter occurs at the membrane level in intact cells, resulting in activation of protein kinase C^{16,17}. Recent evidence concerning the release of bioactive constituents from platelets^{5,8,14,18,19} and neutrophils^{5,6} has linked agonist-induced protein kinase C activation and Ca^{2+} mobilization in a synergistic mechanism^{8,20,21}. Here we present a novel model of

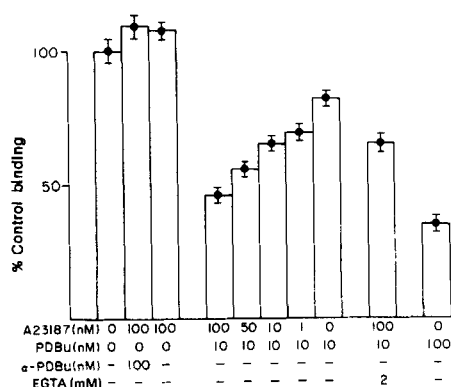


Fig. 2 Effect of dose of A23187 and EGTA on down-regulation of the surface transferrin receptor. Methods as described in Fig. 1 legend. Down-regulation induced by 100 nM phorbol ester in the presence of increasing concentrations of ionophore are shown. Alternatively, 2 mM EGTA was added simultaneously with 100 nM ionophore. Results are expressed as per cent mean control binding $\pm$ s.e.m. of 125 I-ferrotransferrin to cells as described in Fig. 1 legend.

induce terminal differentiation of HL-60 human leukaemic cells in culture¹¹ but also mediate rapid internalization of surface receptors in association with increased phosphorylation of the receptor⁹. Receptor hyperphosphorylation is mediated directly by phorbol ester- and diacylglycerol-activated protein kinase C^{21,22}. Using this system as a model for activated protein kinase C regulation of a specific membrane receptor substrate, a synergistic effect between the 4- β -phorbol-12,13-dibutyrate and the Ca^{2+} ionophore A23187 is observed both on receptor down-regulation and hyperphosphorylation (Fig. 1).

Addition of 100 nM A23187, which is not stimulatory when added alone, augments the effect of 4- β -phorbol-12,13-dibutyrate such that a submaximal concentration (5 nM) produces near-maximal down-regulation (Fig. 1a) and receptor hyperphosphorylation (Fig. 1b). However, addition of ionophore concentrations as high as 1 μM to a maximal phorbol ester concentration of 100 nM does not affect the maximal amplitude of the response. The specificity of this synergistic effect for an active tumour promoter is apparent as the inactive 4- α -phorbol-12,12-dibutyrate, even in the presence of ionophore, failed to have an effect on expression of surface receptor (Fig. 2). The potentiation of the effect of 4- β -phorbol-12,13-dibutyrate is dependent on ionophore concentration and is partially reversed by the addition of 2 mM EGTA (Fig. 2). Ca^{2+} -mediated synergism is similarly observed when the more potent phorbol ester tetradecanoyl phorbol acetate is used instead of 4- β -phorbol-12,13-dibutyrate (data not shown).

Interestingly, the Ca^{2+} -dependent increased activity of phorbol esters described above is associated with an increase in the apparent affinity of ^3H -4- β -phorbol-12,13-dibutyrate binding to whole cells. Results of equilibrium binding of the radioactive ester to cells pretreated with 200 nM A23187 ionophore indicate that a substantial increase in binding occurs at the lower, non-saturating concentrations of 4- β -phorbol-12,13-dibutyrate (3–22 nM) compared with binding in the absence of ionophore (Fig. 3). Scatchard analysis of data demonstrates that the apparent dissociation constant decreases from 18 to 7 nM without any change in the number of binding sites. These data are consistent with the fact that the synergistic effect of Ca^{2+} ionophore is observed only at lower ester concentrations, and suggests that Ca^{2+} mobilization specifically enhances binding at the lower, submaximal concentrations of the active phorbol ester.

Ca^{2+} ionophore also has a marked effect on the kinetics of ester-induced down-regulation of the surface transferrin receptor. Cells were incubated at 37 $^{\circ}\text{C}$ for 5 min with or without ionophore before addition of 100 nM 4- β -phorbol-12,13-dibutyrate. Receptor down-regulation was then estimated at early

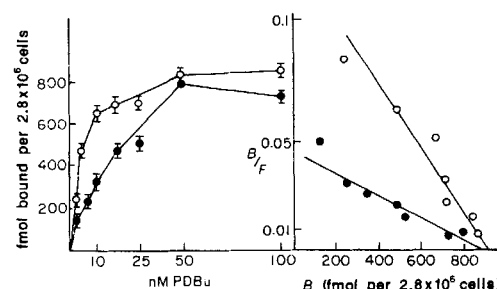


Fig. 3 Effect of A23187 on binding of ^3H -4- β -phorbol-12,13-dibutyrate (PDBu) to cells. Scatchard analysis shows that the dissociation constant in the absence of Ca^{2+} ionophore A23187 (open symbols) was 18 nM. And in the presence of ionophore (closed symbols) was 7 nM. The number of ^3H -ester binding sites per cell in the presence and absence of ionophore is 2.0 and 1.9×10^5 , respectively. The correlation coefficients for least-squares analysis in the presence and absence of ionophore are 0.96 and 0.89 respectively. B/F , ratio of bound to unbound ester.

Methods. Cells were prepared as described in Fig. 1 legend and A23187 was added at 200 nM to cells. Following incubation for 5 min at 37 $^{\circ}\text{C}$, increasing concentrations of unlabelled ester were added to a constant concentration of 3 nM ^3H -4- β -phorbol-12,13-dibutyrate (specific activity 6.5 Ci/mmol⁻¹) and incubation for an additional 40 min. Following incubation, the cells were collected on glass fibre filters (Whatman 6F/C) and washed extensively with ice-cold buffer containing 50 mM Tris-HCl, pH 7.3, 150 mM NaCl. Specific binding of ^3H -ester to cells was determined by scintillation counting in an LKB 1212 Rackbeta liquid scintillation counter. Nonspecific binding of ^3H -ester was obtained by incubating cells in the presence of an excess of unlabelled ester which was subtracted from total binding to give specific binding. Results represent the mean $\pm$ s.e.m. from a representative experiment performed in triplicate. Similar results were obtained each of the four times this experiment was performed. Nonspecific binding accounted for <20% of maximal binding.

(5 min) and late (20 and 45 min) times after addition of ester (Fig. 4). We found a striking difference in the initial rate of receptor down-regulation; in the presence of ionophore, addition of 100 nM ester induced full down-regulation within 5 min. In sharp contrast, the same concentration of ester added in the absence of ionophore induced only a fraction of the maximal response at 5 min than noted at the later time points (Fig. 4). These results suggest that preincubation of cells with ionophore can eliminate any lag time required for recruitment of protein kinase C to the membrane.

Thus, phorbol ester-dependent down-regulation of the transferrin receptor provides a useful model for the study of the synergistic relationship between intracellular Ca^{2+} mobilization and protein kinase C activation. Several relevant conclusions can be drawn from this model: (1) elevating intracellular Ca^{2+} levels, which by itself is ineffective, 'primes' or 'sensitizes' the cells to the action of phorbol ester agonists; (2) higher intracellular Ca^{2+} levels induce an increase in the apparent affinity of HL-60 cells for ^3H -4- β -phorbol-12,13-dibutyrate; (3) the increased affinity is accompanied by a greater potency of phorbol esters with respect to transferrin receptor phosphorylation and down-regulation; (4) the rate of receptor down-regulation, at all effective concentrations of ester, is greatly enhanced by Ca^{2+} ionophore. Therefore, transferrin receptor phosphorylation and down-regulation, which are tightly coupled in this system^{9,21}, seem to be a function of both intracellular Ca^{2+} levels as well as phorbol ester concentration.

Although it has been found that active phorbol esters increase the affinity of protein kinase C for Ca^{2+} to the 10^{-7} M range and that binding of ^3H -ester to purified protein kinase C has an absolute requirement for Ca^{2+} (ref. 8), these findings can be explained in light of results obtained here and from recent studies using an *in vitro* model system (see accompanying paper²³). Consistent with previous findings²⁴, the *in vitro* model

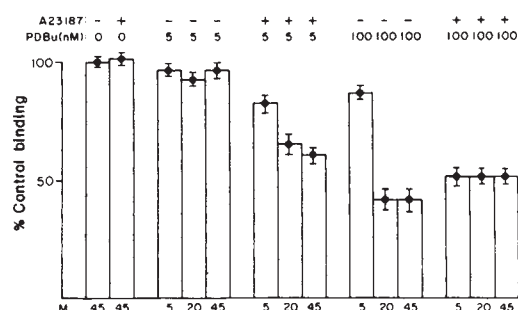


Fig. 4 Effect of A23187 on the rate of PDBu induced down-regulation of the surface transferrin receptor. Additions were made as described in Fig. 1 legend. Cells with added PDBu were incubated for various times from 5 to 45 min as indicated on the abscissa. Following incubation, equilibrium binding with 125 I-ferrotransferrin was performed as described in Fig. 1 legend. Results are expressed as the mean per cent of control binding $\pm$ s.e.m.

shows that phorbol esters also cause a redistribution of protein kinase C to the membrane compartment; however, this effect requires several minutes whereas the Ca^{2+} effect is complete in a few seconds²³. Thus, exposure of HL-60 cells to Ca^{2+} ionophore can effect a rapid redistribution of protein kinase C from the cytoplasm to the plasma membrane. The resulting exposure of the enzyme to membrane phospholipids leads to increased affinity for phorbol esters and a more rapid enzyme activation by these agents. This process is reflected in the increased potency of phorbol esters in stimulating transferrin receptor phosphorylation and down-regulation (Fig. 1), as well as in the increased rate of receptor down-regulation (Fig. 4). Mediation of this membrane process requires activation of protein kinase C^{9,21,25}. Thus, the novel molecular model²³ recognizes that synergism between phorbol ester and Ca^{2+} mobilization occurs at the level of receptor phosphorylation that results from increased ester binding and enzyme activation. This suggests that, at least in part, protein kinase C mediates the effect of Ca^{2+} in this system. The combined effects of Ca^{2+} and phorbol esters (or diacylglycerides^{8,26}) on protein kinase C may also prove relevant for other systems that involve protein kinase C in transmembrane signalling, such as mediation of the action of certain hormones, neurotransmitters and growth factors.

These studies were supported in part by NIH grant IR23 CA37903-01.

Received 11 May; accepted 22 August 1985.

- Wertz, P. W. & Mueller, G. C. *Cancer Res.* **38**, 2900-2904 (1978).
- Diamond, L., O'Brien, T. & Rovera, G. *Life Sci.* **23**, 1979-1988 (1978).
- Rovera, G., Santoli, D. & Damsky, C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2779-2783 (1979).
- Cooper, R. A., Brunwald, A. D. & Kuo, A. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2865-2869 (1982).
- Sha'afi, R. I. et al. *Biochem. biophys. Res. Commun.* **114**, 638-645 (1983).
- White, J. R. et al. *J. biol. Chem.* **259**, 8605-8611 (1984).
- DiVirgilio, F., Lew, D. P. & Pozzan, T. *Nature* **310**, 691-693 (1984).
- Nishizuka, Y. *Nature* **308**, 693-697 (1984).
- May, W. S. Jr, Jacobs, S. & Cuatrecasas, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2016-2020 (1984).
- Cohen, S., Carpenter, G. & King, L. *J. biol. Chem.* **255**, 4834-4842 (1980).
- Cooper, J. A., Bowen-Pope, D. F., Raines, E., Ross, R. & Hunter, T. *Cell* **31**, 263-273 (1982).
- Jacobs, S. J., Sahjoun, N., Saitel, A. S. & Cuatrecasas, P. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6211-6213 (1983).
- Iwashita, S. & Fox, C. F. *J. biol. Chem.* **259**, 2559-2567 (1984).
- Fujikura, T. & Nishizuka, Y. *J. biol. Chem.* **258**, 6701-6704 (1983).
- Rovera, G. et al. *Ann. N.Y. Acad. Sci.* **397**, 211-220 (1982).
- Castagna, M. et al. *J. biol. Chem.* **257**, 7847-7851 (1982).
- Niedel, J. E., Kuhn, L. J. & Vandenberg, G. P. *Proc. natn. Acad. Sci. U.S.A.* **80**, 36-40 (1983).
- Kaibuchi, K., Sano, K., Hoshijima, M., Takai, Y. & Nishizuka, Y. *Cell Calcium* **3**, 323-335 (1982).
- Yamanishi, J. et al. *Biochem. biophys. Res. Commun.* **112**, 778-786 (1983).
- Sutherland, R. et al. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4515-4519 (1981).
- May, W. S., Sahjoun, N., Jacobs, S., Wolf, M. & Cuatrecasas, P. *J. biol. Chem.* **260**, 9419-9426 (1985).
- May, W. S. Jr, Sayhoun, N. E., LeVine, H. III, Jacobs, S. & Cuatrecasas, P. *Clin. Res.* **32**, 498A (1984).
- Wolf, M., LeVine, H. III, May, W. S. Jr, Cuatrecasas, P. & Sahjoun, N. *Nature* **317**, 546-551 (1985).
- Kraft, A. S. & Anderson, W. B. *Nature* **301**, 621-623 (1983).
- Klausner, R. D., Harford, J. & van Renswonder, J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3005-3009 (1984).
- Nishizuka, Y. *Trends biochem. Sci.* **8**, 13-16 (1983).
- Laemmli, U. K. *Nature* **227**, 680-685 (1970).

Site-directed mutagenesis shows that tyrosine 248 of carboxypeptidase A does not play a crucial role in catalysis

Stephen J. Gardell, Charles S. Craik, Donald Hilvert*, Mickey S. Urdea† & William J. Rutter

Hormone Research Laboratory and Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA

The residue Tyr 248 of carboxypeptidase A (CPA) is thought to play a role in catalysis by contributing a proton to the incipient amine anion generated during cleavage of peptide substrates¹. To test this hypothesis we have modified the rat CPA cDNA² by site-directed mutagenesis³ so that the codon for Tyr 248 is replaced by that for Phe. Here, we report the expression of the cDNAs for proCPA and its Tyr-to-Phe variant in yeast via the α -factor system⁴⁻⁶. Following zymogen activation by trypsin, wild-type CPA (CPA-WT) and variant CPA (CPA-Phe 248) were purified to homogeneity and characterized enzymatically. CPA-Phe 248 displays essentially undiminished values for the catalytic constant (k_{cat}) towards various peptide and ester substrates. However, the Michaelis constants (K_m values) of peptide substrates and the inhibition constant (K_i) of the potato carboxypeptidase inhibitor⁷ are increased 6-fold and 70-fold, respectively. These data suggest that the phenolic hydroxyl of Tyr 248 does not act as the requisite general acid catalyst but participates in ligand binding.

CPA catalyses the hydrolysis of aromatic and branched aliphatic amino acids from the carboxy-terminus of peptides and proteins. The amino-acid sequence⁸ and three-dimensional structure of bovine CPA are known⁹. The deduced structure of the enzyme-substrate complex suggests that only two amino acids, Glu 270 and Tyr 248, as well as the essential zinc and water are close enough to the scissile peptide bond to be directly involved in catalysis^{1,10}. Various mechanisms involving these moieties have been proposed¹¹⁻¹⁴. X-ray diffraction analysis of various CPA-ligand complexes shows that Tyr 248 undergoes a substrate-induced conformational change resulting in movement of the phenolic hydroxyl from the surface of the enzyme to within hydrogen-bonding distance of the cleaved peptide bond^{1,10}. Chemical modification of this Tyr residue typically results in decreased peptidase and increased esterase activity¹⁵⁻¹⁷. Tyr 248 was therefore suggested to be the general acid catalyst required for the hydrolysis of peptide bonds. The phenolic hydroxyl was presumed not to be required for cleavage of ester substrates as the alkoxide ion formed during ester hydrolysis is a more stable leaving group than the amide anion, therefore the breakdown of the tetrahedral intermediate need not be dependent on the provision of a proton.

Replacement of Tyr 248 by Phe tests the role of the phenolic hydroxyl in catalysis. The change can be accomplished by site-directed mutagenesis of the CPA cDNA followed by expression of the mutagenized template. We have previously reported the isolation of clones encoding rat CPA from a pancreatic cDNA library². The deduced amino-acid sequence is 78% homologous with the well-studied bovine enzyme and the residues previously implicated in catalysis or ligand binding are conserved between the two species. Therefore, we have carried out the Tyr-to-Phe replacement on the rat CPA cDNA.

Conversion of the Tyr 248 codon (TAT) to Phe (TTT) within the rat CPA cDNA, pCQ1260, was accomplished by oligonucleotide-directed mutagenesis^{18,19}. The expression vectors, pYαCPA-WT and pYαCPA-Phe 248 (Fig. 1a), contain a portion of the α -factor gene²⁰ fused to the cDNAs for proCPA-WT

* Laboratory of Bio-organic Chemistry and Biochemistry, Rockefeller University, New York, New York 10021, USA.

† Chiron Corporation, 4560 Horton Street, Emeryville, California, USA.

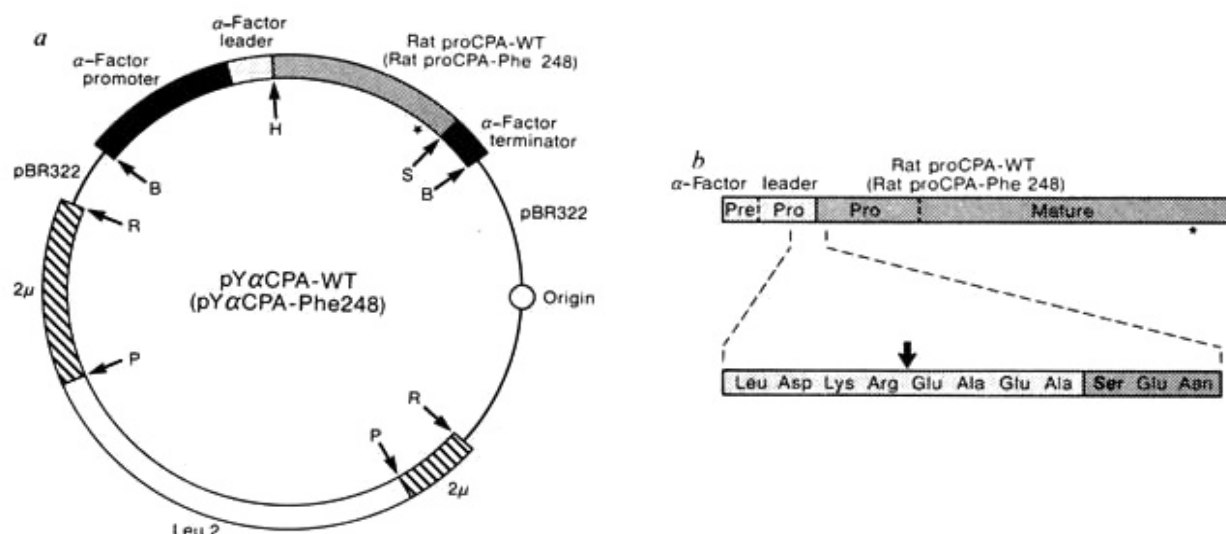


Fig. 1 *a*, Schematic representation of the recombinant plasmids used for expression; *b*, predicted structure of the α -factor leader/proCPA fusion proteins. *a*, Rat CPA cDNA, pCQ1260, was subcloned into M13mp8 (ref. 35) and the single-stranded template, CPA/M13, was isolated. Oligonucleotides 5'AAAGTTCCTCAGAAGCTTCAGCAG3' and 5'GGGATAAGGTTCGACTTTCAGTAGGG3' were synthesized as described elsewhere³⁶ and used to incorporate, respectively, a *Hind*III site at the rat preproCPA junction and a *Sal*I site in the 3' untranslated region by site-directed mutagenesis. The 5' phosphorylated mutagenic primers and the M13 sequencing primer 5'CCCAGTCACGACGTT3' were simultaneously annealed to CPA/M13, extended with T4 DNA polymerase and ligated essentially as described previously¹⁹ except that gene 32 protein was omitted during the extension. Transformation of JM101 and screening for the mutant template, CPA(*Hind*III/*Sal*I)/M13, containing both inserted restriction sites was also carried out as described previously¹⁹. Following plaque purification of CPA(*Hind*III/*Sal*I)/M13, single-stranded DNA was prepared and the entire DNA insert sequenced by the dideoxy-chain terminating method³⁷. Plasmid pAB114 (ref. 5), a pBR322 derivative containing the α -factor gene, was digested with *Hind*III and *Sal*I and ligated to the 1.2-kilobase (kb) *Hind*III/*Sal*I proCPA cDNA fragment from CPA(*Hind*III/*Sal*I)/M13. The resulting vector, p α CPA-WT, contains the in-frame fusion of the α -factor leader (495 base pairs, bp) and rat proCPA sequences under the control of the α -factor promoter and the putative transcription termination signal in the 3' flanking region of the α -factor gene. The 2.2-kb *Bam*HI fragment from p α CPA-WT was introduced at the unique *Bam*HI site of YEpl3 (ref. 38), a yeast plasmid vector which carries the yeast origin of replication and *REP3* from the 2 μ m circle, the yeast *LEU2* gene and a fragment of pBR322 containing the bacterial origin of replication and the ampicillin-resistance gene. The resultant plasmid, pY α CPA-WT, is capable of autonomous replication in both yeast and *E. coli*. The yeast plasmid containing the α -factor leader-proCPA-Phe 248 gene fusion, pY α CPA-Phe 248, was constructed in an identical fashion except for the additional oligonucleotide-directed mutagenesis of CPA/M13 at the position indicated by an asterisk. The mutagenic oligonucleotide used for this purpose, 5'GGCTTGAAAGATTGTGTCGATGAT3', incorporates the desired Tyr 248 to Phe codon change as well as a neutral T $\rightarrow$ C transition. The latter change facilitates screening by increasing the T_m differential between the WT and mutant template with respect to the mutagenic primer. Restriction sites: R, *Eco*RI; B, *Bam*HI; H, *Hind*III; P, *Pst*I; S, *Sal*I. *b*, Arrow indicates the presumed cleavage site recognized by the KEX2 endoprotease²². Ser (in bold letters) has replaced Asn as the first amino acid of the proCPA sequence as a result of the cloning strategy. As the studies on CPA-WT and CPA-Phe 248 involve determination of enzymatic activity following trypsin activation and release of the prosegment, this single amino-acid substitution is inconsequential. An asterisk indicates the approximate position of the Tyr 248 $\rightarrow$ Phe substitution in the corresponding α -factor leader-proCPA-Phe 248 fusion protein.

and proCPA-Phe 248, respectively. The α -factor sequences contribute the promoter; the coding region for both the leader segment and the first spacer peptide; and the transcription termination site. The vectors pY α CPA-WT and pY α CPA-Phe 248 direct the synthesis of hybrid proteins containing the α -factor leader fused to proCPA-WT and proCPA-Phe 248, respectively (Fig. 1b). The spacer peptide Lys-Arg-Glu-Ala-Glu-Ala is situated at the junction of the α -factor precursor and CPA sequences. This site is cleaved by yeast proteinases responsible for processing the α -factor precursor^{21,22}. The proteins released from these α -factor leader fusions appear to be secreted from yeast in an analogous fashion to α -factor⁴⁻⁶. Analysis of media from yeast transformed with either pY α CPA-WT (Fig. 2a, lane 2) or pY α CPA-Phe 248 (lane 3) reveals the accumulation of CPA-immunoreactive species with apparent relative molecular mass (M_r) of 43,000 (43 K) and 34K; these proteins are not present in media from yeast transformed with the parent vector YEpl3 (lane 1). The expression of CPA is dependent on regulation of the linked α -factor promoter, as indicated by the accumulation of CPA-immunoreactive protein in the media from a MAT α (α -factor+) strain harbouring pY α CPA-WT (lane 6) but not an isogenic MAT a/α (α -factor-) strain containing the same plasmid (lane 7).

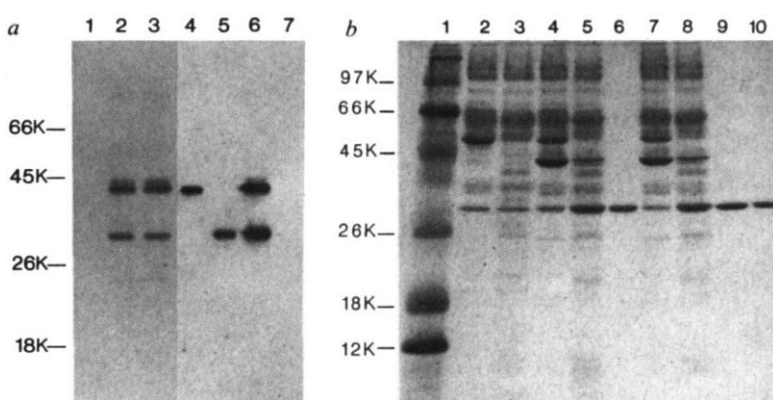
The presence of the 43K and 34K immunoreactive species confirms that processing of the fusion proteins is occurring. The

43K species is smaller than the expected size of rat proCPA (45.5K, predicted from the cDNA sequence) but it has similar mobility to rat proCPA purified from rat pancreas (lane 4). Thus, the discrepancy is apparently the result of aberrant migration during SDS-polyacrylamide gel electrophoresis (PAGE). The minor band migrating just above the 43K species is likely to be an incompletely processed α -factor leader-proCPA fusion protein (lanes 2, 3, 6). The accumulation of the 34K protein (lanes 2, 3, 6) which migrates identically with CPA purified from rat pancreas (lane 5) indicates that yeast can process the procarboxypeptidases. Proteolytic cleavage probably occurs at the normal activation site, an Arg residue at the junction of the pro- and mature segments. A similar trypsin-like cleavage was observed within the β -endorphin peptide following its synthesis via the α -factor-directed expression system (6).

The presence of proCPA and CPA in the media from yeast transformants was also evident by protein staining (Fig. 2b; lane 4; CPA-WT; lane 7, CPA-Phe 248). Trypsin treatment (lanes 5, 8) diminished the 43K band and concomitantly increased the 34K band that co-migrates with rat pancreatic CPA (lane 10). Trypsin-catalysed zymogen activation was substantiated by Western blot analysis (data not shown). Media from control yeast transformed with YEpl3 gave no evidence for the 43K species, but a protein species that co-migrated with rat CPA was present (lane 2). This 34K species is unrelated to CPA as

Fig. 2 a, Western blot analysis of proteins secreted from yeast transformant. Lanes 1-3, yeast strains BJ1994 (*MAT α* , *leu2*, *trp1*, *prb1-1122*, *pep4-3*); lane 6, MH11-1b (*MAT α* , *his3/4*, *trp1*, *leu2⁻*, *ura3-52*) and lane 7, MH11-1b (*MAT α* / α) were transformed by the Li-acetate protocol³⁹ with: lane 1, YEp13; lanes 2, 6, 7, pY α CPA-WT; or lane 3, pY α CPA-Phe248, and leucine prototrophs were selected. Yeast harbouring these plasmids were grown to stationary phase ($A_{600} \sim 7$) in selective media (SD-Leu) containing 50 mM Na succinate (pH 5.5) and 0.1 mM ZnCl₂. Aliquots of media were concentrated by tricarboxylic acid (TCA) (10%) precipitation in the presence of 0.4 mg ml⁻¹ nadeoxycholate before electrophoresis on a 12.5% SDS denaturing gel⁴⁰. The equivalent of 0.1 ml media was added to lanes 1-3, 6 and 7. Proteins were analysed by Western blotting as described elsewhere⁴¹ using rabbit anti-rat CPA antibodies. ProCPA (100 ng) and CPA (100 ng) purified from rat pancreas were applied to lanes 4 and 5, respectively.

Protein relative molecular mass markers are indicated. **b**, Identification of CPA-WT and CPA-Phe 248 in yeast culture media, *in vitro* processing of the proenzyme and purification to homogeneity. Aliquots of media (20 ml) from yeast (BJ1994 *MAT α*) containing YEp13 (lanes 2, 3), pY α CPA-WT (lanes 4, 5) or pY α CPA-Phe 248 (lanes 7, 8) were adjusted to pH 7 and incubated in the absence (lanes 2, 4, 7) or presence of 500 ng trypsin (lanes 3, 5, 8) for 2 h at 37 °C before TCA concentration and SDS-PAGE (12.5%). Each lane represents 2 ml yeast media. Culture media (2 l) from yeast harbouring pY α CPA-WT or pY α CPA-Phe 248 was concentrated by ammonium sulphate precipitation (80%) in the presence of bovine serum albumin (0.5 mg ml⁻¹), resuspended in 20 mM Tris (pH 8.0), 0.5 M NaCl and incubated with 200 μ g trypsin for 1 h at 37 °C to effect proenzyme maturation. Typical yields of either CPA-WT or CPA-Phe 248 present in the medium when cultures are grown to saturation ($A_{600} \sim 7$) were 0.5 mg l⁻¹. Activated CPA-WT or CPA-Phe 248 was dialysed against 20 mM MES (pH 6.0), 0.1 M NaCl, applied to a glycyl-L-tyrosyl-azo-benzyl succinate sepharose affinity column²⁵ and eluted with 20 mM Tris (pH 8.0), 0.5 M NaCl. The final step in the purification involved affinity chromatography on a Sepharose column to which was immobilized the potato carboxypeptidase inhibitor⁷. We used 0.1 M Na₂CO₃, 0.5 M NaCl to elute both CPA-WT and CPA-Phe 248, but whereas the wild-type enzyme required pH 11.4 for efficient elution, the variant enzyme was successfully recovered at pH 10.4. CPA-WT (lane 6) and CPA-Phe 248 (lane 9) were purified to homogeneity as judged by SDS-PAGE; both co-migrated with rat CPA isolated from rat pancreas (lane 10). Staining was with Coomassie brilliant blue. Protein relative molecular mass markers (lane 1) are as indicated.



it does not react with anti-rat CPA antibodies (Fig. 2a, lane 1) and does not accumulate following trypsin treatment (Fig. 2b, lane 3).

Enzymatic hydrolysis of the CPA ester substrate, benzoyl-glycyl- β -L-phenyllactate (Bz-Gly-OPhe), was detected in the media from yeast containing pY α CPA-WT and pY α CPA-Phe 248 but not from yeast containing YEp13. Significantly, the activity was increased proportionately with the relative conversion of proCPA to CPA following treatment with trypsin. Furthermore, the activity was abolished by the chelating agent, o-phenanthroline, which is known to inhibit CPA.

To reconfirm the sequence of the CPA-WT and CPA-Phe 248 templates, pY α CPA-WT and pY α CPA-Phe 248 were isolated by plasmid rescue in *Escherichia coli* from yeast actively synthesizing the putative wild-type and mutant proteins. The sequence of the CPA coding regions revealed that the programmed changes, the A $\rightarrow$ T transversion (yielding CPA-Phe 248) and the neutral T $\rightarrow$ C transition (see legend to Fig. 1a), are the only differences between the templates.

CPA-WT and CPA-Phe 248 were purified by ammonium sulphate precipitation, trypsin activation, affinity chromatography with immobilized benzyl succinic acid²³ and a second affinity chromatography step using the potato carboxypeptidase inhibitor (PCI)⁷. SDS-PAGE showed resulting preparations to be homogeneous (Fig. 2b, lanes 6 and 9, respectively).

The kinetic parameters for hydrolysis of typical peptide and ester substrates by CPA-WT and CPA-Phe 248 are presented in Table I. Comparison of the kinetic constants shows that the k_{cat} values of CPA-WT and CPA-Phe 248 are virtually identical for benzoyl-glycyl-phenylalanine (Bz-Gly-Phe) and Bz-Gly-OPhe. The k_{cat} values for carbobenzoxyglycyl-glycyl-phenylalanine (Cbz-Gly-Gly-Phe) and *p*-chloro-*trans*-cinnamoyl- β -L-phenyllactate (CICPL) were decreased by 2-fold and 4-fold, respectively, as a result of the Tyr 248-to-Phe replacement. In addition, CPA-Phe 248 exhibited increased K_m values for Bz-Gly-Phe (5-fold), Cbz-Gly-Gly-Phe (six-fold) and CICPL (10-fold). The K_m of Bz-Gly-OPhe is not significantly affected by the Tyr 248-to-Phe substitution.

The rate constants exhibited by CPA-WT are identical to those of the enzyme isolated from rat pancreas (data not shown).

Comparison of the kinetic constants of rat CPA for the hydrolysis of Bz-Gly-OPhe with those of the bovine enzyme²⁴ reveals similar k_{cat} and K_m values. Moreover, the k_{cat}/K_m values of bovine CPA for the hydrolysis of Bz-Gly-Phe²⁵, Cbz-Gly-Gly-Phe²⁶ and CICPL²⁷ are of similar magnitude to the rat enzyme. However, both k_{cat} and K_m for these substrates are 5 to 20-fold lower for rat CPA compared with the bovine enzyme.

The effective cleavage of Bz-Gly-OPhe and CICPL by CPA-Phe 248 indicates that the phenolic hydroxyl of Tyr 248 is not required for ester hydrolysis, consistent with previous studies^{15-17,27,28}. However, CPA-Phe 248 and CPA-WT hydrolyse Bz-Gly-Phe and Cbz-Gly-Gly-Phe to a nearly equal degree. Thus, the phenolic hydroxyl moiety is not obligatory for the

Table 1 Kinetic parameters for substrate hydrolysis by CPA-WT and CPA-Phe 248

Substrate	Kinetic parameter	CPA-WT	CPA-Phe 248
Bz-Gly-Phe	k_{cat} (s ⁻¹)	17.7 $\pm$ 0.7	21.2 $\pm$ 0.2
	K_m (μ M)	39.2 $\pm$ 5.1	199 $\pm$ 4
	$10^{-5} k_{cat}/K_m$ (M ⁻¹ s ⁻¹)	4.52	1.07
Cbz-Gly-Gly-Phe	k_{cat}	51.6 $\pm$ 1.9	23.1 $\pm$ 1.6
	K_m	27.1 $\pm$ 3.4	194 $\pm$ 21
	$10^{-5} k_{cat}/K_m$	19.0	1.3
CICPL	k_{cat}	17.5 $\pm$ 0.7	4.13 $\pm$ 0.2
	K_m	26.3 $\pm$ 0.39	252 $\pm$ 27
	$10^{-5} k_{cat}/K_m$	6.65	0.164
Bz-Gly-OPhe	k_{cat}	1194 $\pm$ 62	1138 $\pm$ 41
	K_m	95.6 $\pm$ 7.7	136 $\pm$ 13
	$10^{-5} k_{cat}/K_m$	125	83.8

CPA-WT and CPA-Phe 248 were assayed with Bz-Gly-Phe, Cbz-Gly-Gly-Phe, CICPL and Bz-Gly-OPhe. Hydrolysis was monitored spectrophotometrically at 25 °C in 50 mM Tris HCl, 0.5 M NaCl (pH 7.5) at relatively low substrate concentrations where kinetic anomalies are minimized, and the initial rate data were fitted to Michaelis-Menten kinetics using standard techniques. Protein concentration was determined by amino-acid analysis.

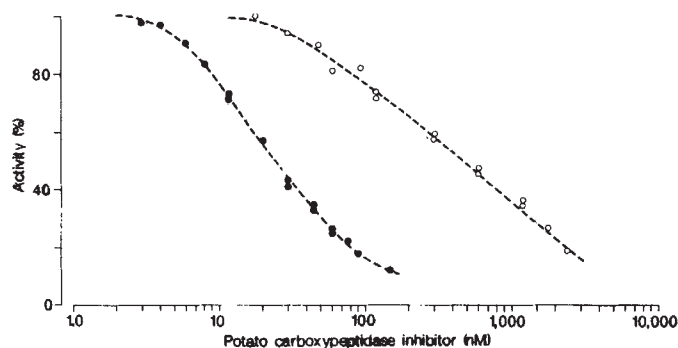


Fig. 3 Inhibition of CPA-WT and CPA-Phe 248 by PCI. Enzyme (0.5 μ g) was preincubated for 15 min with the indicated concentration of PCI in 1 ml 50 mM Tris HCl, 0.5 M NaCl, pH 7.5. Cbz-Gly-Gly-Phe was added to final concentrations of 0.15 mM and 0.5 mM to assay CPA-WT (●) and CPA-Phe 248 (○), respectively. PCI concentration was approximated using the extinction coefficient $E_{280\text{ nm}}^{0.1\%} = 3.0$ (ref. 7).

hydrolysis of peptide substrates, and Tyr 248 is not required to mediate general acid catalysis. If the Tyr 248 hydroxyl acts as a proton donor at all, it can readily be replaced by an alternative group without compromising catalytic activity. The peptidase activity manifested by CPA-Phe 248 is inconsistent with other mechanistic proposals¹⁴. In addition, the formation of the coordination complex involving the phenolic hydroxyl of Tyr 248 and zinc which is observed in the arsanilazo derivative²⁹ cannot be a critical catalytic event.

Other chemical substituents that could fulfil the role of proton donation to the incipient amine anion of the scissile peptide bond include Glu 270 and water^{1,11,12,14}. Based on studies of thermolysin, Monzingo and Matthews³⁰ proposed a mechanism for CPA action in which Glu 270 operates as the requisite proton donor in addition to its classical assignment as a general base. In this mechanism, the proton abstracted by Glu 270 from water during general base catalysis is transferred directly to the amine leaving group of the scissile peptide bond.

The increase in K_m for the cleavage of Bz-Gly-Phe, Cbz-Gly-Gly-Phe and CICPL as a result of the Tyr-to-Phe substitution suggests that the Tyr 248 hydroxyl group participates in substrate binding. However, the effect is not apparent for all substrates; for example, there is little effect on the K_m for the rapidly hydrolysed ester Bz-Gly-OPhe. A role for the phenolic hydroxyl in ligand binding is supported by X-ray diffraction analysis of the glycytyrosine-CPA complex^{1,10} which shows that the hydroxyl group of Tyr 248 can form hydrogen bonds with the NH groups of the scissile and penultimate amide bonds of a peptide substrate.

Because of the effect of the Tyr-to-Phe replacement on the K_m values of most substrates, we tested the ability of PCI to inhibit CPA-WT and CPA-Phe 248. PCI interacts with bovine CPA to form a complex in which residues 37–39 of the inhibitor interact with the active site of CPA³¹. Our experiments show that Tyr 248-to-Phe substitution increases the apparent K_i ~70-fold. The apparent K_i for PCI inhibition of bovine CPA has been reported to be 5 nM (ref. 7). From the data in Fig. 3 and other measurements at varying Cbz-Gly-Gly-Phe concentrations, we determined, using Dixon plots³², that the analogous values for CPA-WT and CPA-Phe 248 are 3 and 200 nM, respectively. The decrease in the apparent affinity of rat CPA for PCI is consistent with the X-ray structure of the bovine CPA-PCI complex; the phenolic hydroxyl of Tyr 248 forms hydrogen bonds to both the amide proton and the carboxylate anion of Val 38 of the inhibitor³¹.

The apparent binding affinity of the complex of rat CPA with Cbz-Gly-Gly-Phe and PCI is decreased by 1.6 and 2.5 kcal mol⁻¹, respectively, due to the Tyr 248-to-Phe replacement. In each case these values are consistent with estimates³³ for the predicted loss of two intermolecular hydrogen bonds.

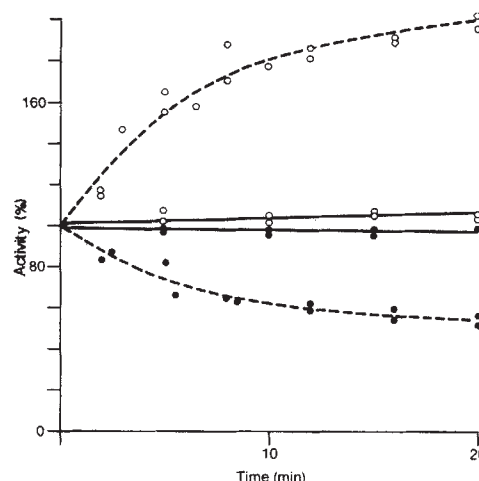


Fig. 4 Effect of tetranitromethane (TNM) on CPA-WT and CPA-Phe 248. Enzyme (30 μ g) was incubated with 10 mM TNM in 50 mM Tris-HCl, 0.5 M NaCl (pH 8.0), at 25 °C. Aliquots were removed at various times and assayed with 1 mM Cbz-Gly-Gly-Phe in 50 mM Tris-HCl, 0.5 M NaCl (pH 8.0), ●—●, CPA-WT; ●—●, CPA-Phe 248; or with 1 mM Bz-Gly-OPhe in 50 mM Tris-HCl, 0.5 M NaCl (pH 8.0); ○—○, CPA-WT; ○—○, CPA-Phe 248.

We also tested CPA-WT and CPA-Phe 248 for the effects of tetranitromethane (TNM), a reagent which selectively nitrates Tyr 248 of bovine CPA³⁴. Under standard conditions of assay (see Fig. 4 legend), nitration of bovine CPA results in decreased activity towards Cbz-Gly-Gly-Phe and increased activity towards Bz-Gly-OPhe¹⁶. TNM treatment of CPA-WT had similar consequences on catalytic activity (Fig. 4). However, TNM treatment had no significant effect on the ability of CPA-Phe 248 to hydrolyse either Cbz-Gly-Gly-Phe or Bz-Gly-OPhe (Fig. 4). This result supports our conclusion that the replacement of Tyr 248 has been accomplished and confirms that the effect of TNM on catalytic activity is a consequence of Tyr 248 modification. The reason for the selective inhibition of peptidase activity by nitration¹⁶ (Fig. 4) or acetylation¹⁵ of Tyr 248 remains to be elucidated. The non-parallel effect on the two classes of substrates cannot be explained by the stringency of the requirement of the phenolic hydroxyl in general acid catalysis. Hence, it seems to reflect mechanistic differences in the manner in which peptide and ester substrates are handled at the active site¹². Further alterations of the amino-acid residues in the active site may help to illuminate the molecular details associated with this interesting but elusive catalytic process.

We thank Dr E. T. Kaiser for helpful discussions; Drs S. Roczniak, E. Fodor, M. Hall, A. Brake, R. Fletterick and C. Largman for their advice and assistance; Dr C. A. Ryan for the gift of potato carboxypeptidase inhibitor; and Ms L. Spector for preparation of this manuscript. This work was supported by NSF grant PCM830610 (to W.J.R.), and NIH grants GM28520 (to W.J.R.) and AM32539 (to E. T. Kaiser). NIH postdoctoral fellowships were awarded to S.J.G. (AM07099), C.S.C. (GM07216) and D.H. (AM07232).

Received 17 May; accepted 20 August 1985.

1. Lipscomb, W. N. *et al. Symp. Biol., Brookhaven* **21**, 24–90 (1968).
2. Quinto, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 31–35 (1982).
3. Hutchison, C. A. *et al. J. biol. Chem.* **253**, 6551–6560 (1978).
4. Emr, S. D., Schekman, R., Flessel, M. C. & Thorner, J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7080–7084 (1983).
5. Brake, A. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 4642–4646 (1984).
6. Bitter, G. A., Chen, K. K., Banks, A. R. & Lai, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5330–5334 (1984).
7. Hass, G. M. & Ryan, C. A. *Meth. Enzym.* **77**, 779–791 (1981).
8. Bradshaw, R. A., Ericsson, L. H., Walsh, K. A. & Neurath, H. *Proc. natn. Acad. Sci. U.S.A.* **63**, 1389–1392 (1969).

9. Rees, D. C., Lewis, M. & Lipscomb, W. N. *J. molec. Biol.* **168**, 367-387 (1983).
10. Rees, D. C. & Lipscomb, W. N. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5455-5459 (1981).
11. Lipscomb, W. N. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3875-3878 (1980).
12. Vallee, B. L., Galdes, A., Auld, D. S. & Riordan, J. F. in *Metal Ions in Biology* Vol. 5 (ed. Spiro, T. G.) 26-75 (Wiley, New York, 1983).
13. Kaiser, E. T. & Kaiser, B. L. *Acc. chem. Res.* **5**, 219-224 (1972).
14. Mock, W. L. & Chen, J. T. *Archs Biochem. Biophys.* **203**, 542-552 (1980).
15. Simpson, R. T., Riordan, J. F. & Vallee, B. L. *Biochemistry* **3**, 616-622 (1963).
16. Riordan, J. F., Sokolovsky, M. & Vallee, B. L. *Biochemistry* **6**, 3609-3617 (1967).
17. Urdea, M. S. & Legg, J. I. *J. biol. Chem.* **254**, 11868-11874 (1979).
18. Zoller, M. J. & Smith, M. *DNA* **3**, 479-488 (1984).
19. Craik, C. S. *et al. Science* **228**, 291-297 (1985).
20. Kurjan, J. & Herskowitz, I. *Cell* **30**, 933-943 (1982).
21. Julius, D., Brake, A., Blau, L., Kunisawa, R. & Thorner, J. *Cell* **37**, 1075-1089 (1984).
22. Julius, D., Schekman, R. & Thorner, J. *Cell* **36**, 309-318 (1984).
23. Cueni, L. B., Bazzone, T. J., Riordan, J. F. & Vallee, B. L. *Analyt. Biochem.* **107**, 341-349 (1980).
24. Whitaker, J. R., Menger, F. & Bender, M. L. *Biochemistry* **5**, 386-392 (1966).
25. Davies, R. C., Riordan, J. F., Auld, D. S. & Vallee, B. L. *Biochemistry* **7**, 1090-1099 (1968).
26. Auld, D. S. & Vallee, B. L. *Biochemistry* **9**, 602-609 (1970).
27. Sub, J. & Kaiser, E. T. *J. Am. chem. Soc.* **92**, 1940-1947 (1976).
28. Glovsky, J., Hall, P. L. & Kaiser, E. T. *Biochem. biophys. Res. Commun.* **47**, 244-247 (1972).
29. Harrison, L. W., Auld, D. S. & Vallee, B. L. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4356-4360 (1975).
30. Monzingo, A. F. & Matthews, B. W. *Biochemistry* **23**, 5724-5729 (1984).
31. Rees, D. C. & Lipscomb, W. N. *J. molec. Biol.* **160**, 475-498 (1982).
32. Dixon, M. *Biochem. J.* **129**, 197-202 (1972).
33. Fersht, A. R. *et al. Nature* **314**, 235-238 (1985).
34. Muszynka, G. & Riordan, J. F. *Biochemistry* **15**, 46-51 (1976).
35. Messing, J. & Vieira, J. *Gene* **19**, 269-276 (1982).
36. Urdea, M. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 7461-7465 (1983).
37. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
38. Broach, J. R. *Meth. Enzym.* **101**, 307-325 (1983).
39. Ito, H., Fukuda, Y., Murata, K. & Kimura, A. *J. Bact.* **153**, 163-168 (1983).
40. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
41. Burnette, W. N. *Analyt. Biochem.* **112**, 195-203 (1981).

A *Drosophila Minute* gene encodes a ribosomal protein

Kritaya Kongsuwan*, Qiang Yu†, Alain Vincent†, Marta C. Frisard†, Michael Rosbash†, Judith A. Lengyel*‡ & John Merriam*§

* Department of Biology, and ‡ Molecular Biology Institute, University of California at Los Angeles, California 90024, USA
† Department of Biology, Brandeis University, Waltham, Massachusetts 02254, USA

Minute genes have long constituted a special problem in *Drosophila* genetics. For at least 50-60 different genes scattered throughout the genome¹⁻⁶, dominant mutations and/or deficiencies have been recognized which result in a common phenotype consisting of short thin bristles, slow development, reduced viability, rough eyes, small body size and etched tergites⁵. Schultz² proposed that the *Minute* loci encode similar but separate functions involved in growth and division common to all cells. Atwood and Ritossa⁷ suggested that *Minute* loci encode components of the protein synthetic machinery, specifically the transfer RNA genes; this now seems unlikely on grounds of both mapping⁸ and mutability studies⁹. More recently, we¹⁰ and others¹¹ suggested that the *Minute* loci are ribosomal protein genes. We report here that transformation with a cloned 3.3-kilobase (kb) region containing the gene encoding the large subunit ribosomal protein 49 (*rp49*) suppresses the dominant phenotypes of *Minute* (*3*)^{99D}, a previously undescribed *Minute* associated with a chromosomal deficiency of the 99D interval¹². This activity is specific to the 99D *Minute* as it does not suppress other *Minute* loci elsewhere in the genome. This result provides direct evidence that the *Minute* locus at the 99D interval encodes the ribosomal protein 49.

rp49 was mapped by *in situ* hybridization to the chromosomal region 99D (ref. 10), near the previously described *Minute* (*3*)¹. Although *M*(*3*)¹ has recently been mapped more precisely to 99B5-9 (ref. 13), distinct from the position of *rp49*, we have used the following genetic approach and identified a previously undescribed *Minute* which maps¹² within the same region as *rp49*. Flies bearing the terminal deficiency chromosome *B81*

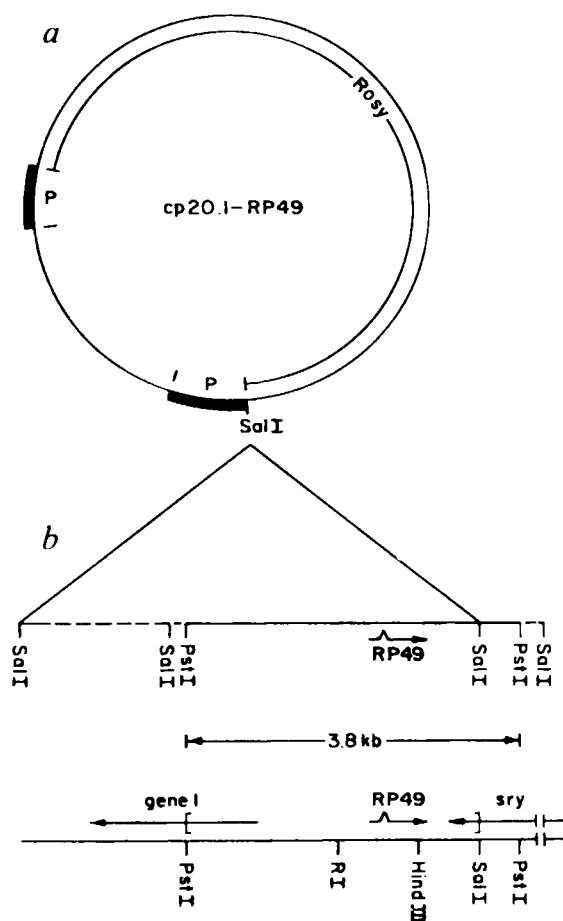


Fig. 1 Construction of cp20.1-RP49 plasmid. *a*, Structure of cp20.1 with *rp49*⁺ gene insert. DNA containing the *rp49*⁺ gene (*b*) was inserted into the unique *SalI* site of cp20.1 (a gift from John Lis, modified from Carnegie 20; ref. 19). *b*, The recombinant phage c25 (ref. 10) was digested to completion with *PstI*. The 3.8-kb *PstI*/*PstI* fragment, containing the *rp49*⁺ gene, was purified from a 1% low melting agarose gel and ligated to *PstI*-digested pUC7 plasmid vector³³. *Escherichia coli* strain JM83 was used for transformation and white colonies were selected on MacConkey plates³⁴. DNA with the entire 3.8-kb *PstI*/*PstI* fragment was digested with *SalI*. A *SalI*/*SalI* DNA fragment containing the *rp49*⁺ gene was purified from a low-melting-point agarose gel, ligated to *SalI*-digested cp20.1 and transformed by standard procedures³⁵ into *E. coli* strain HB101. The structure of cp20.1-RP49 was verified by restriction mapping. The dashed line indicates pUC7 DNA inadvertently included in the fragment cloned into cp20.1. The brackets indicate the *Drosophila* DNA included in the cp20.1-RP49 plasmid. The 3.3-kb DNA fragment inserted into cp20.1 has been well characterized. Much of it has been sequenced¹⁶; it also hybridizes to three *pA*⁺ transcripts (in addition to *rp49* RNA) as it is complementary to the very 3' end of the A and D *sry* transcripts²⁰ and the 5' half of the embryo-specific transcript of gene I (this latter transcript has been mapped by R-looping³⁶; consequently, the fraction of the gene present on the transformed DNA fragment is an estimate). It is formally possible, although unlikely in our view, that one of these gene fragments, or some as yet unidentified region, contains the *Minute*-rescuing activity. This possibility is being investigated.

(*DfB81*)⁶, which lacks the portion of chromosome 3 distal to 99C8, in combination with a duplication (67A) of the tip of the right arm of chromosome 3¹³, have a strong *Minute* phenotype. A series of duplications with different end points was used to show that the *Minute* locus is located within the region 99D2-9 (refs 12, 14). *In situ* hybridization to the same duplication and deficiency chromosomes (data not shown) confirms that the *rp49* gene also lies within the 99D2-9 interval^{14,15} and is uncovered by the synthetic deficiency *B81*-67A (hereafter termed *Dp67A*; *DfB81*).

§ To whom correspondence should be addressed.

Table 1 Genetic analysis of nine *p[rp49⁺]* and three *p[CH8]* transformant lines in crosses segregating *M(3)99D*

Transformant lines	Chromosome location of insert						
	<i>In situ</i> * hybridization	<i>ry</i> ⁺ † marker	<i>Minute</i> ‡ suppressor	1 <i>Ubx</i>	2 <i>Minute</i> non- <i>Ubx</i>	3 non- <i>Minute</i> non- <i>Ubx</i>	4 Intermediate non- <i>Ubx</i>
<i>p[rp49⁺]</i>							
A1	49D	II	II	174	41	52	—
A3	84F	III	III	254	56	50	15
B1	80A	III	III	180	51	51	8
B2	39E	II	—	176	92	0	—
C2	60C	II	II	212	37	62	5
D1	26B, 95C	II, III	II, III	244	34	63	—
D4	26B	II	II	181	51	88	—
F2	80A	III	III	214	27	70	9
F4	45AB	II	II	184	44	47	4
<i>p[CH8]</i>							
814	85AB	III	—	337	171	0	—
829	42B	II	—	301	163	0	—
838	—	III	—	428	224	0	—

The cp20.1-RP49 plasmid construct (Fig. 1) was co-microinjected with p π 25.1 (ref. 32) into *Drosophila* embryos of the genotype *ry*⁵⁰⁶. The microinjected adults were backcrossed to *ry*⁵⁰⁶ flies and their progeny were selected for *ry*⁺ eye colour. Injections and crosses were performed by standard procedures as described elsewhere²⁰. From 387 injected embryos, 62 adults (and 39 fertile adults) emerged, and from these, 7 *ry*⁺-containing vials (A, B, C, D, E, F, G, H) were obtained. Several individuals from each vial (for example, A1, A2, A3, A4) were maintained and propagated as transformed lines. Thus, *p[rp49⁺]* lines with the same letter were derived from the same injected embryo (G0 generation). For *p[CH8]* control lines, three individual *ry*⁺-transformed lines (814, 829, 838) were obtained. For each line, the sum of two crosses involving *Dp67A*; *DfB81*-bearing females to *p[rp49⁺]*- or *p[CH8]*-bearing males is shown. F₁ progeny were scored daily and a fly was identified as a *Minute* if three or more adjacent bristles (macrochaetae) were short and thin (see Fig. 3). Scoring was confined to the dorsal bristles of the head and thorax. The expected ratio of the *Ubx* (class 1):non-*Ubx* (classes 2+3+4) progeny is approximately 2:1 because of the nature of the synthetic deficiency used in these crosses. Only one-quarter (instead of the usual half) of the F₁ progeny received both the *Dp67A* and *DfB81* chromosomes. Those that did not receive the duplication 67A from the mother were heterozygous for the terminal deficiency *B81* and hence were lethal. Some of the transformed *p[rp49⁺]* males were heterozygotes, with respect to the transforming DNA, as evident from the approximately equal proportion of the offspring in the non-*Minute*, non-*Ubx* as in the *Minute*, non-*Ubx* classes. For several lines (C2, D1 and F2) the non-*Minute*, non-*Ubx* progeny were significantly more numerous than the corresponding *Minute* progeny, probably because some of the males used in crosses were homozygous for the insert, and also because some of the inserts had multiple integration sites (for example, D1). See Fig. 2 for a description of the progeny classes 1-3.

* *In situ* hybridization to salivary gland chromosomes was carried out with ³H-labelled RNA probes complementary to the recombinant phage c25 as described previously^{14,15}.

† The chromosomal location of each insert was determined by crosses between females homozygous for *ry*⁴² and males heterozygous for the *p[rp49⁺]* insert(s) and bearing dominantly marked second and third balancer chromosomes (none of the transformed lines tested showed sex-linked inheritance).

‡ Mapping of the insert(s) using the *M*⁺ phenotype was done by crossing males heterozygous for the *p[rp49⁺]* insert and the dominant markers *Curly* on the second chromosome and *Stubble* on the third chromosome to *Dp67A*; *DfB81*-bearing females. The *Minute* suppressor was assigned to the second and/or third chromosome based on the *M*⁺ or *Minute* phenotype of the appropriate segregants.

If the 99D *Minute* is caused by the absence of one copy of the *rp49* gene, the introduction of an additional copy of the gene by P-element-mediated DNA transformation should suppress the *Minute* phenotype. A plasmid containing the *rp49* gene was constructed for this rescue experiment. The entire *rp49*⁺ coding region, as well as ~0.6 kb of downstream and ~2.1 kb of upstream DNA^{16,17} (the region indicated in brackets in Fig. 1b), were introduced into the cp20.1 *ry* transforming vector^{18,19} at the *SalI* site, to give the plasmid construct cp20.1-RP49 (Fig. 1). Homozygous *ry* embryos were injected with cp20.1-RP49 plasmid DNA, and transformants, designated *p[rp49⁺]*, were selected on the basis of their ability to express the *ry*⁺ phenotype (Fig. 1, 2, Table 1).

To test for suppression of the *Minute* phenotype by the cloned ribosomal protein gene, the 99D *Minute* deficiency and the *p[rp49⁺]* insert were combined in the same flies by crossing *ry*⁺ males from the *p[rp49⁺]* transformant lines to *Dp67A*; *DfB81* females. The results for nine representative *p[rp49⁺]* lines are shown in Table 1. For most of these crosses, three progeny classes were observed: (1) Non-*Minute*, *Ubx* flies derived from eggs carrying the balancer chromosome¹² *DpUbx^LMcp^R*; (2) *Minute*, non-*Ubx* flies, which carried *Dp67A*; *DfB81*, but did not receive the *ry*⁺ marked *p[rp49⁺]* insert as confirmed by test crosses of these F₁ males to *ry* females (data not shown); (3) non-*Minute*, non-*Ubx* flies which did not receive the *DpUbx^LMcp^R* chromosome and must, therefore, carry *Dp67A*; *DfB81*, but were wild type (*M*⁺) in appearance, judging by the size and length of the macrochaetae on the head and dorsal thorax (Fig. 3). That this latter class received the *ry*⁺-

marked *p[rp49⁺]* insert from their fathers was confirmed by test crosses of the F₁ males to *ry* females (data not shown). The presence of the *Dp67A*; *DfB81*-bearing chromosomes in these apparently *M*⁺ flies was also verified by outcrossing them to wild-type (Oregon R) flies; in all cases, *Minute* progeny were recovered among the offspring (data not shown). In control experiments, flies transformed with the construct cp20.1-CH8 (ref. 20) (containing an 8-kb *HindIII* fragment including the

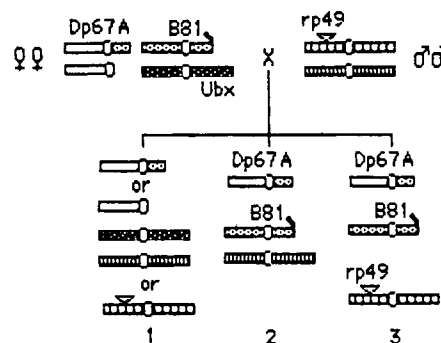


Fig. 2 Progeny of crosses. Because of its strong *Minute* effect, *Dp67A*; *DfB81* in combination with most third chromosome balancers is poorly viable and sterile. It was therefore maintained over chromosome¹² *DpUbx^LMcp^R*, which carries a duplication of the 99D region to suppress the *Minute*, and an inversion of the 89-99 interval to suppress crossing over; these flies are *Ubx* and *M*⁺ in phenotype.

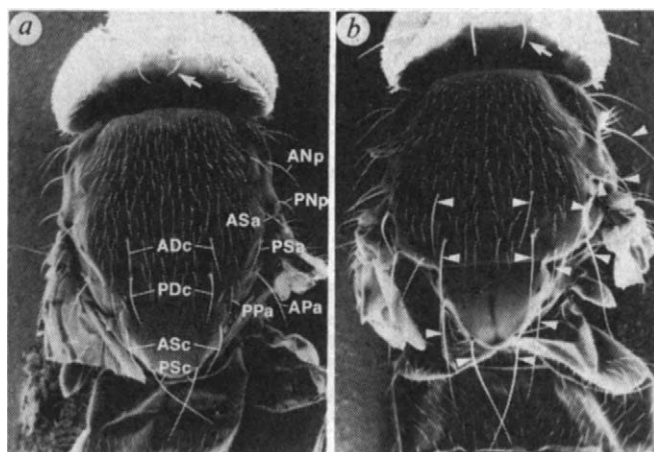


Fig. 3 $p[rp49^+]$ suppresses the short and thin bristle phenotype of $M(3)99D$. *a*, Scanning electron micrograph³⁷ of the dorsal thorax of a $Dp67A; DfB81/+$ *Minute* female showing the characteristic small and thin macrochaetae. A, PNp = notopleurals, A, PSa = supraalar, A, PDc = dorsocentrals, A, PPa = postalar and A, PSc = scutellars where A = anterior and P = posterior. *b*, Micrograph of the dorsal thorax of a $Dp67A; DfB81/p[rp49^+]$ female taken at the same magnification ($\times 62.9$) as *a*, showing normal-size macrochaetae. Corresponding macrochaetae labelled in *a* are indicated in *b* by arrowheads. Note also the substantial difference in the length of the postvertical bristles on the head (arrows).

adjacent *serendipity* (*sry*) genes^{17,20}; Fig. 1) did not suppress the 99D *Minute* phenotype in similar crosses (Table 1), nor did the $p[rp49^+]$ -transformed flies suppress 15 *Minute* genes mapping at other loci (data not shown; the *Minute* loci shown not to be suppressed by the $p[rp49^+]$ insert are $M(1)n$, $M(2)173$, $M(2)H^{55}$, $DfM(2)c^{33a}$, $M(2)m^{56}$, $M(2)S2^3$, $M(3)be^{36e}$, $M(2)Z$, $M(3)1$, $DfM(3)f$, $M(3)h^y$, $DfM(3)S31$, $M(3)S32$, $M(3)w$ and $DfM(4)^{62F}$).

As a further test of the notion that suppression of the 99D *Minute* phenotype is caused by the transformed $rp49^+$ sequences, we mapped the $p[rp49^+]$ inserts independently by *in situ* hybridization and by genetic crosses which followed the segregation of the ry^+ marker and the *Minute* suppressor (data not shown). The results of these mapping experiments for eight lines showed that in every case the chromosome sites of the $p[rp49^+]$ DNA, the ry^+ marker and the *Minute* suppressor are in agreement (Table 1), as expected if the ry^+ -marked $p[rp49^+]$ acts as the *Minute* suppressor.

Not all the $p[rp49^+]$ inserts suppress, or suppress to an equal extent. In line B2, $p[rp49^+]$ did not suppress the 99D *Minute* phenotype. Perhaps the insert lost, or did not receive, an active $p[rp49^+]$, despite the presence of the insert as detected by *in situ* hybridization with phage c25 DNA (Table 1). In this line, the $p[rp49^+]$ insertion site was at 39E (near the histone genes at 39D-E; ref. 21), which is near the chromocentre (Table 1). It is possible that some position effect at this location results in a low level of transcription of the $p[rp49^+]$ insert. In five other lines (A3, B1, C2, F2 and F4), although the $p[rp49^+]$ insert completely suppresses the *Minute* phenotype in most flies, a few flies with an intermediate phenotype are found (Table 1). Test crosses of the intermediate-*Minute* males to *rosy* females confirm the presence of the $p[rp49^+]$ insert (data not shown). Although these data suggest that the level of $p[rp49^+]$ expression is not the same in all transformants, all transformants are ry^+ . This is not surprising because wild-type eye pigmentation requires <1% of the normal level of xanthine dehydrogenase (XDH)^{22,23}. We conclude that the *Minute* phenotype associated with the deletion of the 99D interval is specifically suppressed by the $p[rp49^+]$ insert but that this expression may be somewhat variable and position dependent, as shown for other transduced genes^{24,25}.

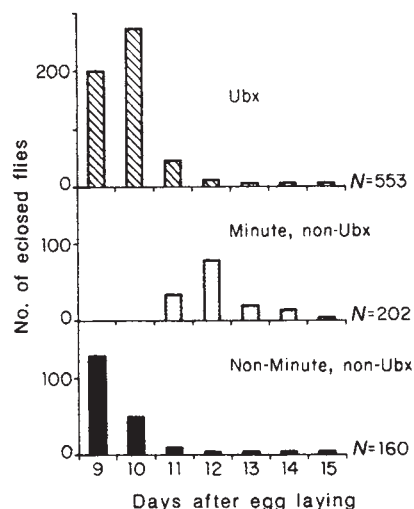


Fig. 4 Developmental time (from egg to adult) for the three classes of progeny resulting from the cross between $M(3)99D$ and $p[rp49^+]$ flies (Fig. 2, Table 1). $Dp67A; DrB81/Dp(3)Ubx^L/Mcp^R$ females were mated to $p[rp49^+]$ males of line A1. After 2 days together, the cross was allowed to lay eggs in active yeast-supplemented medium for 12 h. Development was at 25 °C. After emergence, the adult progeny were collected twice daily. To eliminate any bias in scoring, six replicate cultures were independently scored by two observers. The results did not differ significantly between the observers and were therefore summed for analysis. *N*, number of flies examined in each class. The non-*Minute* $DpUbx^L/Mcp^R$ flies required an average of 9.9 ± 0.4 days to emerge after egg laying. The *Minute* $Dp67A; DfB81$ flies showed prolonged development, requiring on average 12.2 ± 0.4 days after egg laying to emerge as adults; this is significantly longer than that of the control sibs carrying the *Ubx* duplication (*G*-test, $P < 0.01$). The non-*Minute*, non-*Ubx* flies, which carried the $p[rp49^+]$ insert, exhibited a normal developmental time, requiring an average of 9.6 ± 0.5 days to emerge as adults. This was not significantly different from that of the control sibs ($P > 0.6$).

In addition to suppressing the thin-bristle *Minute* phenotype, the $p[rp49^+]$ insert also restores the normal developmental rate of the 99D *Minute*-bearing progeny (Fig. 4). By 10 days after egg laying, 83% of the non-*Minute*, non-*Ubx* flies had emerged as adults, compared with only 2% of the *Minute* flies. The $p[rp49^+]$ inserts thus suppress not only the *Minute* bristle phenotype but also the slow development syndrome.

The $p[rp49^+]$ insert also suppresses a previously undescribed maternal effect on embryonic segmentation pattern associated with the 99D *Minute*. The 99D *Minute* females, when mated with wild-type males, laid eggs which were, on average, smaller than normal (data not shown). Also, some 10–15% of the unhatched eggs died as mature embryos and exhibited a specific cuticular defect; the abdominal segments posterior to A3 were often fused and partially deleted (Fig. 5). This abdominal phenotype is not observed in embryos from $Dp67A; DfB81$ mothers bearing $p[rp49^+]$ (Fig. 5); in addition, eggs from these mothers are of normal size (data not shown). Thus, the $p[rp49^+]$ insert specifically suppresses the maternal effects of $M(3)99D$.

The multiplicity and phenotypic similarity of the various *Drosophila Minute* loci, as well as their effect on cell division rate²⁶, allow the prediction that many of them encode ribosomal proteins. Presumably, mutations at *Minute* loci lead to the *Minute* phenotype by lowering the rate of ribosome assembly and, hence, the rate of protein synthesis²⁷. A substantial variation in messenger RNA levels for different ribosomal proteins²⁸, in the absence of dosage compensation²⁷, may explain the observations that there are fewer *Minute* loci than eukaryotic ribosomal proteins^{6,29}, and also that there are both weak and strong *Minute* mutations². For genes whose mRNA levels are almost twice what is required to sustain a normal growth rate, a heterozygous

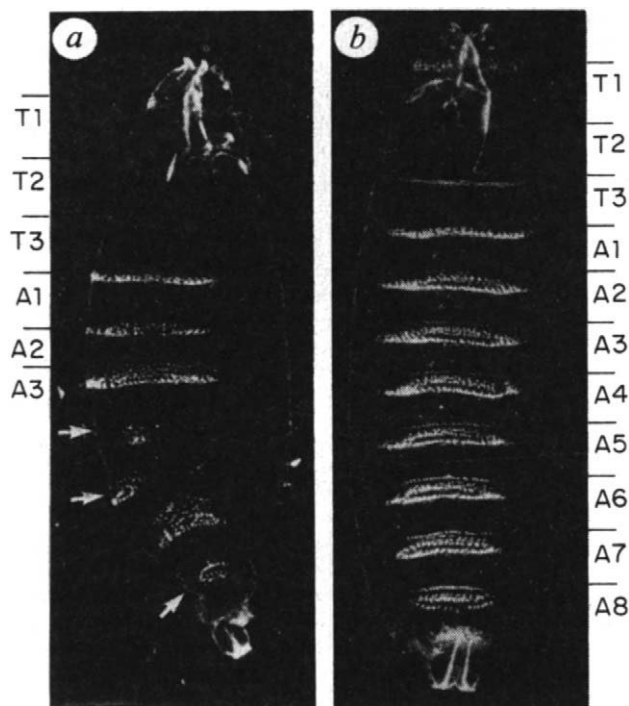


Fig. 5 Ventral cuticular patterns of embryos obtained from crosses of wild-type males with *Dp67A; DfB81 Minute* females (a), and with *Dp76A; DfB81/p[rp49⁺]* non-*Minute* females (b). Embryos were fixed and mounted as described previously³⁸. For a detailed description of cuticular pattern, see ref. 39. A, abdominal segment; T, thoracic segment. a, A characteristic segmentation defect; the abdominal segments posterior to A3 are fused and partially deleted, leaving a remnant of the fused denticle bands. This phenotype is suppressed by the *p[rp49⁺]* insert since unhatched embryos with the segmentation defect are not observed from the *p[rp49⁺]*-bearing females.

deficiency would be expected to have an undetectable, or only a mild phenotype. For genes at the other extreme, that is, those whose mRNA levels are only barely in excess of what is required (almost rate-limiting), a heterozygous deficiency would be expected to have a severe phenotype. Presumably *M(3)99D* falls in this latter class, as the heterozygous deficiency is a strong *Minute* and is not fully rescued in all of the transformed lines, that is, phenotypic rescue by the transduced gene is sensitive to modest position-effect modulation of gene expression.

A primary effect of *Minute* mutations on ribosome assembly would also explain two interesting observations of Schultz²: Three wild-type doses of a particular *Minute* gene do not suppress a mutation in a different *Minute* gene; and the phenotype of double or triple *Minute* lines is no more severe than that of a line carrying only the most extreme single *Minute* mutation of the combination. We propose that the mutation giving rise to the lowest ribosomal protein mRNA level is rate limiting for ribosome assembly. The presence of other less severe mutations will have no apparent effect. In this light, Schultz's observations suggest that many *Minute* mutations are effectively deletions which result in reduced levels of ribosomal proteins, and thus retard the rate of ribosome assembly. They may be true deletions or, equally likely, some of them may be missense or nonsense mutations^{30,31} which encode mutant ribosomal proteins that cannot assemble.

We thank R. MacIntyre, E. B. Lewis and L. Craymer for fly stocks, and J. A. Lake and J. C. Hall for helpful comments on the manuscript. This work was supported by NSF grant PCM 84-03539 to J.A.L. and J.R.M. and by NIH grants GM23549 and P01GM33205 to M.R.

Received 9 May; accepted 5 August 1985.

1. Bridges, C. B. & Morgan, T. H. *Carnegie Inst. Wash. Publ.* 327, 251 (1923).
2. Schultz, J. *Genetics* 14, 366-419 (1929).

3. Dunn, L. C. & Coyne, J. *Biol. Zentr.* 55, 385-389 (1935).
4. Dunn, L. C. & Mossige, J. C. *Hereditas* 23, 70-90 (1937).
5. Lindsley, D. L. & Grell, E. H. *Carnegie Inst. Wash. Publ.* 627, (1968).
6. Lindsley, D. L. *et al. Genetics* 71, 157-184 (1972).
7. Ritossa, F. M., Atwood, K. C. Spiegelman, S. *Genetics* 54, 819-834 (1966).
8. Kubli, E. in *Genetic Map 1984. A Compilation of Linkage and Restriction Maps of Genetically Studied Organisms* Vol. 3 (ed. O'Brien, S. J.) 300-303 (Cold Spring Harbor Laboratory, New York, 1984).
9. Huang, S. L. & Baker, B. S. *Mutat. Res.* 34, 407-414 (1976).
10. Vaslet, C. A. O'Connell, P., Izquierdo, M. & Rosbash, M. *Nature* 285, 674-676 (1980).
11. Burns, D. K., Stark, B. C., Macklin, M. D. & Chooi, W. Y. *Molec. cell Biol.* 4, 2643-2652 (1984).
12. Kongsuwan, K., Dellavalle, R. & Merriam, J. *Genetics* (in the press).
13. Frisardi, M. C. & MacIntyre, R. *Molec. gen. Genet.* 197, 403-413 (1984).
14. Lengyel, J. A. *et al. Belle W. Baruch Symp.* 15 (1985).
15. Roark, M., Mahoney, P., Graham, M. & Lengyel, J. A. *Dev. Biol.* 109, 476-488 (1985).
16. O'Connell, P. & Rosbash, M. *Nucleic Acids Res.* 12, 5495-5513 (1984).
17. Vincent, A., O'Connell, P., Gray, M. R. & Rosbash, M. *EMBO J.* 3, 1003-1013 (1984).
18. Rubin, G. M. & Spradling, A. C. *Science* 218, 348-353 (1982).
19. Rubin, G. M., Spradling, A. C. *Nucleic Acids Res.* 11, 6341-6351 (1983).
20. Vincent, A., Colot, H. V. & Rosbash, M. *J. molec. Biol.* (in the press).
21. Pardue, M. L., Kedes, L. H., Weinberg, E. S. & Birnstiel, M. L. *Chromosoma* 63, 135-151 (1977).
22. Chovnick, A., Gelbart, W. & McCarron, M. *Cell* 11, 1-10 (1977).
23. Spradling, A. C. & Rubin, G. M. *Cell* 34, 47-57 (1983).
24. Scholnick, S. B., Morgan, B. A. & Hirsch, J. *Cell* 34, 37-45 (1983).
25. Hazelrigg, T., Levis, R. & Rubin, G. M. *Cell* 36, 469-481 (1984).
26. Morata, G. & Ripoll, P. *Dev. Biol.* 42, 211-221 (1975).
27. Abovich, N. & Rosbash, M. *Molec. cell Biol.* 4, 1871-1879 (1984).
28. Meyuhas, O. & Perry, R. P. *Gene* 10, 113-129 (1980).
29. Chooi, W. Y., Sabatini, L. M., Macklin, M. & Fraser, D. *Biochemistry* 19, 1425-1433 (1980).
30. Sinclair, D. A. R., Suzuki, D. T. & Grigliatti, T. A. *Genetics* 97, 581-606 (1981).
31. Sinclair, D. A. R., Grigliatti, T. A. & Kaufman, T. C. *Genet. Res.* 43, 257-275 (1984).
32. Karess, R. E. & Rubin, G. M. *Cell* 38, 135-146 (1984).
33. Viera, J. & Messing, J. *Gene* 19, 259-268 (1982).
34. Miller, J. H. *Experiments in Molecular Genetics*, 49 (Cold Spring Harbor Laboratory, New York, 1972).
35. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning. A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
36. Wong, Y. C., O'Connell, P., Rosbash, M. & Elgin, S. C. R. *Nucleic Acids Res.* 9, 6749-6752 (1981).
37. Hodgkin, N. M. & Bryant, P. J. in *The Genetics and Biology of Drosophila* Vol. 2c (eds Ashburner, M. & Wright, T. R. F.) 337-358 (Academic, New York, 1978).
38. Van der Meer, J. *Drosoph. Inf. Serv.* 52, 160 (1977).
39. Lohs-Schardin, M., Cremer, C. & Nusslein-Volhard, C. *Dev. Biol.* 73, 239-255 (1979).

Corrigenda

Primary structure and genomic organization of the histidine-rich protein of the malaria parasite *Plasmodium lophurae*

J. V. Ravetch, R. Feder, A. Pavlovic & G. Blobel
Nature 312, 616-620 (1984)

Since publication of this article, sequence analysis of the genomic clone 8A using the chemical degradation technique of Maxam and Gilbert has revealed that the M13 clones used to generate the sequence published in Fig. 3 has undergone rearrangement due to the repetitive nature of the sequence cloned. A deletion of 90 nucleotides had occurred in the repeat region 938-1087. In contrast to the 5 published repeats of the sequence APHHHHHHHH, 8 repeats are found in the genomic clone by Maxam-Gilbert sequencing. These are encoded as 3 copies of the sequence GCTCCACACCATCATCACCACCATCACCACCAT followed by 5 copies of the sequence GCTCCACACCATCATCACCACCATCACCACCAT. Recombination in the M13 clone had occurred between repeat 2 and 5 generating the 3 repeat deletion. Similarly, in the region 1178-1474, encoding 10 copies of the sequence DAHHHHHHHH in the M13 subclones, only 6 copies are found by sequence analysis of the genomic clone using chemical degradation sequencing techniques representing an insertion of 120 bp in this region. Evidence for the presence of an insertion and deletion has been independently obtained from sequence analysis of a partial cDNA clone corresponding to the region 926-1,648 (Irving *et al. Molec. biochem. Parasit.*, in the press). These results reflect the potential difficulty in using subcloning into single-stranded phages to obtain DNA sequence data on highly repetitive genes of this type.

The effect of climate on long-term changes in the crustacean zooplankton biomass of Lake Windermere, UK

D. G. George & G. P. Harris
Nature 316, 536-539 (1985)

In line 5 of the first paragraph on p. 537 the area of the south basin should read 6.78 km² and the last word on line 18 should read copepodites.

Oncogene chromosome breakpoints and *Alu* sequences

It has recently been shown^{1,2} that the human leukaemia-associated 'Philadelphia chromosome' results from a reciprocal translocation which fuses the *c-abl* gene on chromosome 9 with the *bcr* gene on chromosome 22. However, the exact position of the breakpoint in one of the recently published sequences¹ needs clarification, because repeated sequences of the *Alu* dispersed repeat family³ are present. This breakpoint, like some others, may have resulted from illegitimate recombination within an *Alu* sequence.

In one of the cases reported by Heisterkamp *et al.*¹ (0319129), the authors noted homology to *Alu*; an *Alu* unit on chromosome 9 has recombined with a partly similar but misaligned sequence on chromosome 22 (see Fig. 1). In the other case (02120185), the presence of *Alu* sequences was overlooked and makes the identification of the breakpoint ambiguous. As shown in Fig. 1, the translocation chromosome (9q+) again has an *Alu* unit recombined with a non-homologous sequence from chromosome 22. This *Alu* unit lacks the first 103 nucleotides of the standard *Alu* sequence, and its truncated 5' end is joined to the chromosome 22 sequence with a 2-nucleotide overlap (nucleotides 109–110 of Fig. 3b of ref. 1; vertical bar in our Fig. 1).

The data presented do not make it clear whether this is the actual 9;22 junction. The 'chromosome 9' sequence aligned by the authors appears to be a different *Alu* sequence—in fact, two 3'-half-*Alu* sequences joined by a long A-rich tract—with numerous (13/81) differences from the 9q+ *Alu* sequence. If the latter is indeed from chromosome 9, it is probably derived from a different *Alu* unit, which could be located in chromosome 9 clones by identifying unique chromosome 9 sequences at the other end of the *Alu* sequence. Alternatively, it may be a 5'-truncated *Alu* unit inserted into chromosome 22 sequen-

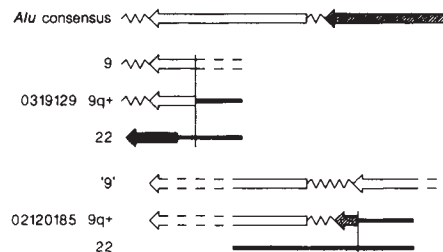


Fig. 1 Disposition of *Alu* sequences in the sequences of Heisterkamp *et al.*¹. In each case the *Alu* (+)-strand is homologous to the (-)-strand of the published sequences. The 300-nucleotide *Alu* consensus (top line)³ consists of two homologous halves, each followed by an A-rich tract which is indicated by zigzags. The boxed sequence indicates a partly *Alu*-related sequence noted by Heisterkamp *et al.*¹.

ces on 9q+; this would be an unlikely coincidence, but 5'-truncated insertions do occur⁴.

The fact that one and probably both of the 9;22 translocations occurred within *Alu* sequences is interesting in itself. Four separate thalassaemias have been caused by illegitimate recombination within *Alu* sequences⁵; so has one case of familial hypercholesterolaemia⁶. All these had breakpoints at different points in the *Alu* sequence. Because *Alu* sequences constitute 3–5% of the human genome, these occurrences could be coincidental⁵, but it is possible that *Alu* sequences are hot spots for illegitimate recombination.

JOHN ROGERS

*MRC Laboratory of Molecular Biology,
Hills Road,
Cambridge CB2 2QH, UK*

1. Heisterkamp, N. *et al. Nature* **315**, 758–761 (1985).
2. Shitvelman, E. *et al. Nature* **315**, 550–554 (1985).
3. Schmid, C. W. & Jelinek, W. R. *Science* **216**, 1065–1070 (1982).
4. Rogers, J. H. *Int. Rev. Cytol.* **93**, 187–279 (1985).
5. Vanin, E. F. *et al. Cell* **35**, 701–709 (1983).
6. Lehrman, M. A. *et al. Science* **227**, 140–145 (1985).

HEISTERKAMP AND GROFFEN REPLY—We agree with Rogers that the positioning of the breakpoint in chronic myelogenous leukaemia (CML) patient 02120185 is subject to discussion. Although we discussed some aspects of the consequences of the Ph¹ translocation in this patient¹, we would like to elaborate on this issue. We have shown previously² that the 9q+ fragment of this patient represents a chromosome 9;22 chimaeric DNA fragment. The normal chromosome 9 DNA sequences were subsequently isolated using the chromosome 9 part of the 9q+ as a probe, thus demonstrating that the chromosome 9 and 9q+ sequences must form a set. Subsequently, we have localized the 'breakpoint' on the normal chromosome 9 fragment by comparative restriction enzyme analysis of the normal chromosome 9 and 9q+ fragments. On a nucleotide level, the situation is less clear. As we have indicated, 13 nucleotides within an 81-base-pair (bp) stretch are different between the 9 and 9q+ chromosome; moreover, 27 bp on the 9q+ chromosome do not originate from the chromosome 22 or chromosome 9 sequences shown.

Although the sequences shown¹ must represent the actual 9q+ sequences present in patient 02120185, we cannot with certainty define the mechanism by which this configuration was attained; we favour the explanation that on translocation of the chromosome 22 sequences to chromosome 9, the resulting 9q+ chromosome was unstable, and a deletion, probably between *Alu* sequences, occurred. The boxed 9q+ sequences would then represent the original 9q+ sequences while sequences more to the 5' would originate from sequences located more towards the centromere of the breakpoint on chromo-

some 9. Nonetheless, we cannot exclude the possibility suggested by Rogers that a 5'-truncated *Alu* unit was inserted into the chromosome 22 sequences on the 9q+ chromosome.

It is also noteworthy that part of the *Alu* sequences on chromosome 9 of 02120185 (14 bp 3' of the breakpoint) are homologous with the *Alu* sequences of chromosome 9 in patient 0319129: of 42 nucleotides in 02120185, 36 can be found surrounding the breakpoint in patient 0319129. This suggests that the breakpoints on chromosome 9 fall within similar types of *Alu* repeat sequences. However, on chromosome 22, breakpoints and *Alu* repeats do not seem to coincide; although *Alu* sequences can be found in the sequences around the breakpoint on chromosome 22 in patient 0319129, similar sequences seem to be absent in patient 02120185. Moreover, regions of *bcr* containing numerous breakpoints for different CML patients have been used as repeat-free probes² in Southern hybridizations, indicating that these regions lack *Alu* repetitive sequences. Therefore, *Alu* sequences may be hot spots for recombination, but not all illegitimate recombination events in CML seem to occur within repetitive sequences.

NORA HEISTERKAMP

JOHN GROFFEN

*Oncogene Science, Inc.,
222 Station Plaza North,
Mineola, New York 11501, USA*

1. Heisterkamp, N., Stam, K., Groffen, J., de Klein, A. & Grosveld, G. *Nature* **315**, 758–761 (1985).
2. Groffen, J. *et al. Cell* **36**, 93–99 (1984).

The age of (a tiny part of) the Australian continent

SCHÄRER and Allègre¹, using isotope dilution analysis of single grains and fragments of zircon, recently failed to substantiate a report by Froude *et al.*², based on the ion microprobe SHRIMP³, of zircons older than 4,100 Myr from Mt Narryer, Western Australia. In their discussion, Schärer and Allègre disparaged the ion microprobe results on various grounds, although they did not actually state that they were wrong. Here, we wish to give a simple explanation for Schärer and Allègre's null observation, to correct an evident misreading of Froude *et al.*, and to advise that the ages of the >4,100-Myr-old zircons are unchanged following reanalysis by SHRIMP and that several more such grains have been discovered.

Schärer and Allègre did not discuss the statistical implications of the very low abundance of the >4,100-Myr-old grains (hereafter termed the 'old' grains) found by Froude *et al.*, which was 4 in a total of 102 zircons analysed at that time. Suppose for the moment that this abundance is exactly right, so that the chance of selecting a single young grain will be 0.96; it follows that the chance of selecting 32 young grains in succession, which Schärer

and Allègre have done, is $(0.96)^{32}$, that is, 0.27. Although 27% is not a very high probability, it is nevertheless far too high to be disregarded. Schärer and Allègre have not analysed a sufficiently large number of zircons to decide whether there are any old grains in the original Froude *et al.* concentrate or not. All that can be concluded from their results alone is that the abundance of old grains is less than about 9%, which would give a probability of successively selecting 32 young grains of <0.05 . If we now accept that Froude *et al.*'s abundance estimate of 0.04 for the old grains has an uncertainty of the order of 0.02, it becomes obvious that many more grains than 32 must be analysed before the existence of the old grains could be excluded with any confidence. Schärer and Allègre admit that grain selection as the reason for their lack of old grains 'cannot be completely eliminated', but that single sentence is not an adequate assessment of the statistical issue.

We regret that Schärer and Allègre were apparently misled by Fig. 3 of Froude *et al.* into believing that all four old zircons had the same appearance as Froude *et al.*'s grain 34. Grain 34 alone is unique: morphologically, the other old grains cannot be distinguished from the rest of the population. For this reason, Froude *et al.* analysed at random all except the very turbid zircons rather than a prior selection according to this or that visual criterion. Schärer and Allègre did make a prior non-random selection. It seems possible, therefore, that they may have unwittingly biased their grain selection against the old zircons.

Our current estimate for the abundance of old zircons in the same concentrate used by Schärer and Allègre is 5 in 260. Each of the original four old grains has been reanalysed by ion probe (with good agreement), and one more has been discovered in a total of 158 new zircons analysed. At this lower observed abundance, the chance of missing an old grain in 32 successive selections is slightly more than 50%. On this basis, it is no surprise that Schärer and Allègre did not find one.

Schärer and Allègre perceive differences in discordancy patterns between their data

and those of Froude *et al.* We do not agree that there are any significant differences between the two sets of analyses, as long as the greater experimental uncertainty in measuring Pb/U by ion probe is taken into account. In fact, we consider that Schärer and Allègre have confirmed Froude *et al.*'s results for the 3,750 Myr-old and younger zircons. The data for all 20 of Froude *et al.*'s young zircons fall along and extend the discordancy band shown in Schärer and Allègre's Fig. 2a between 3,750 Myr and 1,400 Myr. Those that fall above Concordia mainly represent valid excursions along a correlated $y-x$ error locus due to uncertainty in determining Pb/U, which is greater for the ion probe (see Table 1 of Froude *et al.*²²) than for isotope dilution analyses to which conventional zircon geochronologists are accustomed. A few analyses above Concordia for the old zircons belong there because of local metamorphic redistribution of Pb, as documented recently by Williams *et al.*⁴. It is true that such 'reverse discordance' is rarely if ever found in conventional work. This probably reflects the difference in sampling scale between ion probe analyses, which typically consume a few nanograms of zircon, and conventional analyses which at best consume a few micrograms⁵ and typically many milligrams, thereby averaging out any fine-scale variation in Pb/U.

Schärer and Allègre also consider that the contrast between their consistently discordant ages for the young zircons and the 'apparently concordant' ages of the old grains measured by Froude *et al.* presents a difficulty. The facts are that most of Froude *et al.*'s ion probe analyses of the old grains are slightly below Concordia, and two at most are above it. We concede that if Froude *et al.* had a systematic error of a few per cent too high in their measurements of Pb/U, the old grains would appear to be more concordant than they really are. However, Froude *et al.*'s results for the Isua zircons do not show a major effect of this sort. On the other hand, there is the following central question which Schärer and Allègre do not address: if the ion probe ages of $>4,100$ Myr are wrong, how can the high and very consistent values of $^{207}\text{Pb}/^{206}\text{Pb}$ for these particular samples be explained? The corrections for common Pb have negligible effects, the observed precision of the ratios is high and due mainly to counting statistics, there are no known isobaric interferences under ^{207}Pb , and the discrimination between Pb isotopes during sputtering (in contrast to Pb^+ and U^+) is negligible. Why do Froude *et al.* measure the correct $^{207}\text{Pb}/^{206}\text{Pb}$ for the Isua zircons also?

Thus, we believe that Schärer and Allègre have simply been unlucky in failing to find any $>4,100$ -Myr-old zircons, and that this is understandable when their low abundance is considered. There is no question of the reality of the old grains. Two more $>4,100$ -Myr-old zircons have

been found recently in a second quartzite horizon at Mt Narryer⁶.

Finally, the preliminary interpretations of the young zircon data from the Mt Narryer quartzites given by Froude *et al.* and followed by Schärer and Allègre are now superseded by conclusions in new papers (refs 6, 7 and P. D. Kinny *et al.*, in preparation), which include discussion of the zircon age discordancy patterns, the timing of deposition and metamorphism of the quartzite, and the provenance of the (young) detrital zircons.

W. COMPSTON
D. O. FROUDE
T. R. IRELAND
P. D. KINNY
I. S. WILLIAMS
I. R. WILLIAMS
J. S. MYERS

Research School of Earth Sciences,
Australian National University,
GPO Box 4,
Canberra, ACT, Australia

1. Schärer, U. & Allègre, C. J. *Nature* **315**, 52-55 (1985).
2. Froude, D. O. *et al.* *Nature* **304**, 616-618 (1983).
3. Compston, W., Williams, I. S. & Meyer, C. J. *geophys. Res. Suppl.* **89**, B525-B534 (1984).
4. Williams, I. S. *et al.* *Contr. Miner. Petrol.* **88**, 322-327 (1984).
5. Schärer, U. & Allègre, C. J. *Nature* **295**, 585-587 (1982).
6. Froude, D. O. thesis, Aust. National Univ. (1985).
7. Wijbrans, J. *et al.* *Contr. Miner. Petrol.* (submitted).

SCHÄRER AND ALLÈGRE REPLY—The comment of our Australian colleagues is not surprising, but does not provide any new information, except one speculation which is incorrect. Zircon samples have not been selected according to any special criterion such as one which would eliminate old grains. Our 39 analyses, representing 32 grains, include zircons of all the different crystal types distinguishable in this sample. The probability calculations given above would rather reinforce our cautious but clear conclusions.

We are more interested in the truth than in promoting any peculiar technique. The only solution to the dilemma would be a re-examination of the $>4,000$ -Myr-old zircons by isotope dilution, and an extensive study by both techniques on identical grains from populations with simple and well defined histories.

We thus take this opportunity to reiterate our offer to analyse any grain that they can send us, even just one, two or three. Furthermore, we would be grateful if the Isua ion probe data, to which the authors refer so frequently, could be published soon.

U. SCHÄRER

Royal Ontario Museum,
Mineralogy and Geology Department,
100 Queen's Park,
Toronto, Ontario,
Canada M5S 2C6

C. J. ALLEGRE

Laboratoire de Géochimie et Cosmochimie,
Universités de Paris, 6 et 7,
4, Place Jussieu,
75230 Paris Cedex 05,
France

Matters Arising

Matters Arising is meant as a vehicle for comment and discussion about papers that appear in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted. Neither contribution nor reply (if one is necessary) should be longer than 500 words and the briefest of replies, to the effect that a point is taken, should be considered.

Physics

Quarks, Gluons and Lattices. By M. CRUETZ. Cambridge University Press: 1985. Pp.169. ISBN 0-521-31535-2. \$12.95, £7.95.

Spectroscopy of Condensed Media: Dynamics of Molecular Interactions. By C.H. WANG. Academic: 1985. Pp.356. ISBN 0-12-734780-1. \$74, £59.50.

Springer Proceedings in Physics, Vol. 3: Nonlinear Phenomena in Physics. F. CLARO (ed.). Springer-Verlag: 1985. Pp.439. ISBN 3-540-15273-3. DM 98.

Springer Series in Electrophysics, Vol. 13: Physics of Highly Charged Ions. By R.K. JANEV, L.P. PRESNYAKOV and V.P. SHEVELKO. Springer-Verlag: 1985. Pp.330. ISBN 3-540-12559-0/0-387-12559-0. DM 148.

Springer Series in Electrophysics, Vol. 18: Planar Circuits for Microwaves and Lightwaves. By T. OKOSHI. Springer-Verlag: 1985. Pp.199. ISBN 3-540-13853-6/0-387-13853-6. DM 118.

Springer Series in Optical Sciences, Vol. 47: Tunable Solid State Lasers. P. HAMMERLING, A.B. BUDGOR and A. PINTO (eds). Springer-Verlag: 1985. Pp.203. ISBN 3-540-15135-4/0-387-15135-4. DM 83.

Topics in Applied Physics, Vol. 58: Hot-Electron Transport in Semiconductors. L. REGGIANI (ed.). Springer-Verlag: 1985. Pp.275. ISBN 3-540-13321-6/0-387-13321-6. DM 124.

Turbulent Shear Flows 4. Selected Papers from the Fourth International Symposium on Turbulent Shear Flows. L.J.S. BRADBURY et al. (eds). Springer-Verlag: 1985. Pp.395. ISBN 3-540-13744-0/0-387-13744-0. DM 168.

Chemistry

Analytical Solution Calorimetry. J. KEITH GRIME (ed.). Wiley: 1985. Pp.401. ISBN 0-471-86942-2. £69.40.

Chemistry of Hydrocarbon Combustion. By D.J. JUCKNALL. Chapman & Hall: 1985. Pp.415. ISBN 0-412-26110-3. £39.50.

Handbook on the Physics and Chemistry of Rare Earths, Vol. 7. K.A. GSHNEIDNER Jr and L. EYRING (eds). North Holland: 1984. Pp.579. ISBN 0-444-86851-8. Dfl. 335.

Impact Evaluation and Regulation of Chemicals, Vol. 2: Impact of Regulations — Improvement of Methods. F. HOMBURGER (ed.). Karger: 1985. Pp.326. ISBN 3-8055-3942-8. SwFr. 218, DM 262, \$93.

Synthetic Reagents, Vol. 6: Chloramine-T, Hydrogen Peroxide, Polyphosphoric Acid. J.S. PIZEY (ed.). Ellis Horwood: 1985. Pp.438. ISBN 0-85312-739-5. £45.

Technology

High-Speed Fiber Spinning: Science and Engineering Aspects. ANDRZEJ ZIABICKI and HIROMICHI KAWAI (eds). Wiley: 1985. Pp.586. ISBN 0-471-89792-2. Np.

Integrated Technology for Parallel Image Processing. S. LEVIALDI (ed.). Academic: 1985. Pp.236. ISBN 0-12-444820-8. \$39.50, £39.50.

Interventional Ultrasound. H.H. HOLM and J. KVIST KRISTENSEN (eds). Munksgaard: 1985. Pp.186. ISBN 87-16-09776-9. Dkr. 300.

Optical Fiber Communications, Vol. 1: Fiber Fabrication. TINGYE LI (ed.). Academic: 1985. Pp.363. ISBN 0-12-447301-6. \$54, £54.

Third European Congress on Biotechnology: Proceedings, Vol. IV. FEDERATION OF BIOTECHNOLOGY. VCH: 1985. Pp.592. ISBN 0-89573-524-5. Np.

Computer Sciences

Computer-Based Automation. JULIUS T. TOU (ed.). Plenum: 1985. Pp.652. ISBN 0-306-41903-3. Np.

Engines of the Mind: A History of the Computer. By JOEL SHURKIN. W.W. Norton: 1985. Pp.352. ISBN 0-393-01804-0. £18.50, \$22.50.

6th European Conference on Electrotechnics: Eurocon 84. Computers in Communication and Control. Peter Peregrinus Ltd: 1984. Pp.424. ISBN 0-86341-029-4. £54.

Introduction to Logic Programming. By CHRISTOPHER JOHN HOGGER. Academic: 1984. Pp.278. ISBN 0-12-352090-8. \$46, £35.50.

System Design with Microprocessors, 2nd Edn. By D. ZISSOS. Academic: 1985. Pp.191. ISBN 0-12-781750-6. \$19.50, £12.

Earth Sciences

Geology of the Northern Philippine Sea: Geological Results of the GDP Cruises of Japan. TSUNEMASA SHIKI (ed.). Tokai University Press, Shinjuku Tokai Building, 3-27-4 Shinjuku, Shinjuku-Ku, Tokyo 160, Japan: 1985. Pp.288. ISBN 4-486-00876-6. Np.

Geomorphology Texts, Vol. 8: Limestone Geomorphology. By STEPHEN TRUDGILL. Longman: 1985. Pp.196. ISBN 0-582-30011-8. £12.95.

Hydrological Forecasting. M.G. ANDERSON and T.P. BURT (eds). Wiley: 1985. Pp.604. ISBN 0-471-90614-X. £36.50.

Lecture Notes on Coastal and Estuarine Studies, Vol. 11. Observations on the Ecology and Biology of Western Cape Cod Bay, Massachusetts. JOHN D. DAVIS and DANIEL MERRIMAN (eds). Springer-Verlag: 1984. Pp.289. ISBN 3-540-96084-8/0-387-96084-8. DM 58.

Lecture Notes in Earth Sciences, Vol. 1: Sedimentary and Evolutionary Cycles. ULF BAYER and ADOLF SEILACHER (eds). Springer-Verlag: 1985. Pp.465. ISBN 3-540-13982-6/0-387-13982-6. DM 65.

Longman Dictionary of Geography: Human and Physical. By AUDREY N. CLARK. Longman: 1985. Pp.724. ISBN 0-582-35261-4. £25.

Applied Biological Sciences

Cognition, Stress and Aging. JAMES E. BIRREN and JUDY LIVINGSTON (eds). Prentice/Hall: 1985. Pp.212. ISBN 0-13-139825-3. £36.85.

Lithium: Current Applications in Science, Medicine, and Technology. RICHARDO O. BACH (ed.). Wiley: 1985. Pp.422. ISBN 0-471-80073-2. £22.50.

Multiple Sclerosis: Experimental and Clinical Aspects. L. SCHEINBERG and C.S. RAINE (eds). The New York Academy of Sciences: 1985. Pp.518. Pbk ISBN 0-89766-267-9. Pbk \$120.

Neuro Pharmacology '85. H. KELEMEN, K. MAGYAR and E.S. VIZI (eds). Akademiai Kiado, P.O.B. 24, H-1363 Budapest: 1985. Pp.351. ISBN 963-05-4057-6. £31.25.

Neurophysiological Correlates of Mental Disorders. B. SALETY, C. PERRIS and D. KEMALI (eds). Karger: 1985. Pp.111. ISBN 3-8055-3969-X. SwFr 85, DM 102, \$36.25.

Occupational and Industrial Hygiene: Concepts and Methods. Advances in Modern Environmental Toxicology, Vol. 8. NURTON A. ESMEN and MYRON A. MEHLMAN (eds). Princeton Scientific: 1984. Pp.240. ISBN 0-911131-09-4. £68.

Prevention of Physical and Mental Congenital Defects, 2 Vols. MAURICE MAROIS (eds.). Alan R. Liss: 1985. Vol. 1: pp.492, ISBN 0-8451-0184-6, £75; Vol. 2: pp.435, ISBN 0-8451-0185-4, £73.

Principal Aspects of Clinical Nutrition. J.C. SOMOGYI and R. WENGER (eds). Karger: 1985. Pp.132. ISBN 3-8055-3950-9. DM 135, \$67.75.

Progress in Clinical and Biological Research, Vol. 172A: Molecular Basis of Cancer. Part A: Macromolecular Structure, Carcinogens, and Oncogenes. ROBERT REIN (ed.). Alan R. Liss: 1985. Pp.576. ISBN 0-8451-0186-2. £59.

Renal Effects of Petroleum Hydrocarbons. Advances in Modern Environmental Toxicology, Vol. 7. M.A. MEHLMAN et al. Princeton Scientific: 1984. Pp.284. ISBN 0-911131-08-6. £62.85.

Reshaping Life: Key Issues in Genetic Engineering. By G.I.V. NOSSAL. Cambridge University Press: 1985. Pp.158. Pbk ISBN 0-521-31603-0. Pbk £6.95, \$11.95.

Spheroids in Cancer Research: Recent Results in Cancer Research, Vol. 95. H. ACKER et al. (eds). Springer-Verlag: 1984. Pp.183. ISBN 3-540-13691-6/0-387-13691-6. DM 98.

Thymic Factor Therapy. N.A. BYROM and J.R. HOBBS (eds). Raven Press: 1985. Pp.441. ISBN 0-88167-041-3. \$33.50.

Toxicity Testing: New Approaches and Applications in Human Risk Assessment. A.P. Li (Editor-in-Chief). Raven: 1985. Pp.280. ISBN 0-88167-083-9. \$50.

The Value of Preventive Medicine. CIBA FOUNDATION SYMPOSIUM 110. Pitman: 1985. Pp.258. ISBN 0-272-79784-7. \$35.

Psychology

Practical Approaches to Alcoholism Psychotherapy, 2nd Edn. SHELDON ZIMBERG, JOHN WALLACE and SHEILA B. BLUME (eds). Plenum: 1985. Pp.406. ISBN 0-306-41762-6. Np.

Surviving Schizophrenia: A Family Manual. By E. FULLER TORREY. Harper & Row: 1985. Pp.306. ISBN 0-06-091217-0. £7.50.

Sociology

Manufacturing Industry and Economics in the SADCC Countries: Energy Environment and Development in Africa, Vol.5. By RICHARD PEET. The Beijer Institute and the Scandinavian Institute of African Studies: 1984. Pp.119. ISBN 91-7106-233-5. SEK 75.

Self-Organization and Management of Social Systems. H. ULRICH and G.J.B. PROBST (eds). Springer-Verlag: 1984. Pp.155. ISBN 3-540-13459-X/0-387-13459-X. DM 65, \$22.90.

Social Science Research Methods. By BRUCE A. CHADWICK, HOWARD M. BAHR and STAN L. ALBRECHT. Prentice/Hall: 1984. Pp.454. ISBN 0-13-818336-8. £31.

Wood, Energy and Households: Perspectives on Rural Kenya. Energy Environment and Development in Africa, Vol. 6. CAROLYN BARNAS (ed.). The Beijer Institute and the Scandinavian Institute of African Studies: 1984. Pp.213. ISBN 91-7106-235-1. SEK 75.

Archaeology

Contributions to Palaeobotany. M.W. DICK and D. EDWARDS (eds). Academic: 1984. Pp.225. Pbk ISBN 0-12-215120-8. Pbk \$16, £9.60.

Data Processing in Archaeology. By J.D. RICHARDS and N.S. RYAND. Cambridge University Press: 1985. Pp.232. ISBN 0-521-25769-7. £31.50.

General

Current Topics in Chinese Science. Vol 1, Chemistry; Vol. 2, Biology. Gordon & Breach: 1984. Vol. 1: pp.563, ISBN 0-677-40385-2. Vol. 2: pp.496, ISBN 0-677-40415-8. Vol. 1, \$68; Vol. 2, \$60.

Cypress Swamps. KATHERINE CARTER EWEL and HOWARD T. ODUM (eds). University of Florida Press: 1985. Pp.472. ISBN 0-8130-0714-3. \$25.

Decision and Control in Uncertain Resource Systems. By MARC MANGEL. Academic: 1985. Pp.255. ISBN 0-12-468720-2. \$39.50, £30.50.

Developments in Expert Systems. M.J. COOMBS (ed.). Academic: 1984. Pp.253. ISBN 0-12-187580-6. \$15, £9.50.

Discourses in Reading and Linguistics. SHEILA J. WHITE and VIRGINIA TELLER (eds). The New York Academy of Sciences: 1984. Pp.225. ISBN 0-89766-261-X. Np.

Drink and Drinking Problems. By CLIVE EASTMAN. Langman: 1985. Pp.208. Pbk ISBN 0-582-29608-0. Pbk £3.95.

Energy: Crisis or Opportunity? An Introduction to Energy Studies. By DIANA SCHUMACHER. MACMILLAN, London: 1985. Pp.335. ISBN 0-333-27191-2. Hbk £25. Pbk £12.

Enumeration and Design. DAVID M. JACKSON and SCOTT A. VANSTONE (eds). Academic: 1984. Pp.317. ISBN 0-12-379120-0. \$39.50, £30.50.

Environmental Dispute Resolution. By LAWRENCE S. BACOW and MICHAEL WHEELER. Plenum: 1984. Pp.372. ISBN 0-306-41594-1. \$29.50.

Environmental Politics and Policy. By WALTER A. ROSENBAUM. Congressional Quarterly, 1414 22nd Street, NW, Washington: 1985. Pp.328. Pbk ISBN 0-87187-306-0. Pbk \$10.50.

Canada's research manpower

from Richard Pearson

A boom in jobs in research and development is forecast in Canada, but these will be in the new technologies and industry. Traditional academic careers will remain hard to get.

THE Canadian government, like many others, is committed to significantly increasing the level of research and development in the country and the adoption of new technology. Such progress, however, is seen to be conditional on the availability of highly skilled personnel; even if the technology is imported, highly skilled personnel are still needed to apply it effectively.

In 1979 the Canadian government set a target to increase expenditure on research and development to reach 1.5 per cent of the gross national product (GNP) within five years. In the event this was not achieved and the total expenditure in 1984, \$C5,000 million, represented only 1.25 per cent of GNP. The growth in research manpower fell similarly short of target, the number increasing by only 8,000 to total the equivalent of 28,000 full-time scientists and engineers in research in 1984. Over half were employed in industry, which has shown the fastest growth in research personnel over the past decade, while one-quarter were in government employment and 16 per cent in the universities (Table 1).

The latest assessment of the adequacy of Canada's research manpower resources up to 1990, in a report issued by the

or very slow economic growth". The imbalances were found to be most serious for applied scientists and engineers, new researchers with research qualifications, and staff in areas such as computer science and biotechnology, where shortages were expected under any scenario.

Annual demand for new PhD students could rise as high as 3,000 per year modelled on scenarios of high growth and rising expenditure on research and development, or be as low as 600 per year with no growth in research or GNP. The corresponding figures for MSc students range from 1,300 to 3,800, and both sets of figures showed marked differences according to discipline. As there was believed to be a plentiful supply of BSc students, detailed projections were not made.

Traditionally, the majority of new researchers have come from the ranks of newly qualifying graduates and postgraduates. Last year there were over 16,000 full-time postgraduates studying relevant disciplines in Canadian universities, a dramatic increase on the 11,000 five years earlier. However, one-third of the growth was accounted for by foreign students, with nearly all of the increase in the applied sciences and engineering coming from overseas students. In 1984 they accounted for over half the students in these subjects and more than one in three studying physical sciences and mathematics, a pattern not dissimilar to that found in the United States and Britain. In Canada, were it not for this growth, many university departments would have had problems maintaining their graduate programmes and research.

Looking forward, the number of postgraduates is already largely determined for the period to 1990, given the long lead times necessary to adjust university programmes. The number of Canadian postgraduates awarded MScs is expected to rise from 2,500 in 1984 to 3,900 in 1990, but less than half are expected to enter employment directly. The corresponding figures for PhDs are 670, rising to 994, although all of the PhDs are expected to seek work, postdoctoral work being regarded as employment. The average annual supply of postgraduate researchers entering the labour market over the next five years will be around 1,950 at the MSc level and 850 PhDs, but under half are in the shortage subject areas of applied sciences and engineering (Table 2).

The council believes that the shortages

that will exist under any realistic scenario can be reduced only by a combination of factors. These include raising the productivity of researchers (measured as less researchers per unit of research expenditure) and increasing the inflow of suitably

Table 2 Average annual supply of postgraduates 1984-90

	MScs	PhDs
Applied sciences and engineering	906	198
Physical sciences and mathematics	418	295
Agricultural and biological sciences	314	211
Health professions	314	139
Total	1,952	842

Source: *Research Talent in the Natural Sciences and Engineering, Supply and Demand Projections to 1990* (NSERCC, 1985).

qualified staff from jobs not connected with research and giving them appropriate retraining, although this latter option is not thought likely to yield significant results if US experience is repeated. As the growth in research posts is largely centred on industrial needs, another route is for companies to upgrade the research skills of existing staff, particularly those at first-degree level. If this were to be done in collaboration with university programmes, the universities themselves would be beneficiaries of industrial experience. While immigration is not seen to be a solution, given worldwide shortages, a more productive opportunity could be to make it easier for overseas students to stay on in Canada by relaxing immigration criteria.

Within the education system, one major constraint on increasing the supply of new postgraduates is that there are not enough qualified students in the pipeline were the postgraduate funding programme to be increased. There is thus a longer-term need to improve science teaching in schools and to encourage more students, particularly women, to study relevant subjects.

Demand in key disciplines such as computer science, biotechnology and specialist engineering is expected to grow particularly fast, and much of the growth in research jobs will be in industry rather than academia or government. As such, potential entrants to research and development will have to have an education that permits flexibility to be able to respond to changing needs and will have to look beyond the traditional career routes to benefit from any research boom. □

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.

Table 1 Scientists and engineers in research and development*

Year	Government	Industrial	University
1975	5,990	8,320	3,840
1976	6,160	9,020	3,960
1977	6,250	9,720	4,030
1978	6,430	10,520	3,750
1979	6,410	11,310	4,140
1980	6,440	12,980	4,210
1981	6,610	14,640	4,160
1982	7,320	15,900	4,390
1983	7,467	16,165	4,387
1984	7,572	16,093	4,416

*Full-time equivalents including BSc, MSc, PhD (natural sciences and engineering only).

National Sciences and Engineering Research Council of Canada, raises serious concern about future manpower constraints.

Using a model which can test a range of possible assumptions, such as differing rates of economic growth, levels of investment in research, 'productivity' of researchers, and the effective working life of researchers, the council's report argues that the supply of research talent "would be inadequate to meet the requirements, even with a fairly modest growth in R & D expenditures. It would be adequate only under less desirable conditions such as no